LLM-Powered Cognitive Assistants for Knowledge Sharing among Factory Operators

Kernan Freire, S.

DOI
10.4233/uuid:733e10e7-890c-48a7-86d3-fa782ffd65c8

Publication date
2025

Document Version
Final published version

Citation (APA)
Kernan Freire, S. (2025). *LLM-Powered Cognitive Assistants for Knowledge Sharing among Factory Operators*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:733e10e7-890c-48a7-86d3-fa782ffd65c8

LLM-Powered Cognitive Assistants for Knowledge Sharing among Factory Operators

LLM-Powered Cognitive Assistants for Knowledge Sharing among Factory Operators

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday, 15 January 2025 at 15:00 o'clock

by

Samuel KERNAN FREIRE

Master of Science in Industrial Design Engineering,
Technical University of Delft,
The Netherlands
born in London, The United Kingdom

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. A. Bozzon	Delft University of Technology, *promotor*
Dr. E. Niforatos	Delft University of Technology, *copromotor*

Independent members:

Prof. dr. ir. G.J.P.M. Houben	Delft University of Technology
Prof. dr. A. Schimidt	Ludwig-Maximilians-Universität München, Germany
Prof. dr. P. Markopoulos	Eindhoven University of Technology
Prof. dr. G. Mentzas	National Technical University of Athens, Greece

Other members:

Prof. dr. P.S. César Garcia	Delft University of Technology

Keywords: Knowledge Sharing, Conversational AI, Socio-technical Systems, Research through Design, Large Language Models, Human Computer Interaction, Knowledge Management

Printed by: Ridderprint

Cover by: S. Kernan Freire

An electronic copy of this dissertation is available at https://repository.tudelft.nl/.

Contents

Summary

Modern factories show little resemblance to the assembly lines from a century ago. Nowadays, a single human might be responsible for setting up, operating, and maintaining a complex chain of machines. This shift has made the work of factory operators more cognitively demanding, requiring constant monitoring and problem-solving. Key to tackling these challenges is the effective collaboration of human operators in exchanging ideas and knowledge, from high-level problem-solving strategies to solutions for emerging issues. Yet, the heightened productivity requirements and small shifts mean fewer opportunities to share this knowledge face-to-face and reporting practices are deprioritized. Thus, the manufacturing industry faces a knowledge management crisis.

This dissertation investigates the integration of conversational AI assistants into manufacturing settings to facilitate knowledge sharing among factory operators. Capitalizing on recent advancements in Natural Language Processing (NLP), particularly in Large Language Models (LLMs), this research investigates the designing and evaluation of conversational AI tools that efficiently capture and share human knowledge on the factory floor while addressing operator needs and concerns. The introduction of conversational AI assistants for knowledge sharing—referred to as cognitive assistants (CA) in this work—in factory environments promises significant benefits but comes with numerous challenges.

Through a research-through-design approach, we conduct ethnographic studies, prototype development, and empirical research. By designing and deploying several functional prototypes in real-world production settings, this work investigates how to effectively design cognitive assistants aligned with human and organizational needs to facilitate knowledge sharing in factories. We iteratively developed several CA prototypes to extract useful information from (unstructured) natural language, dynamically update knowledge bases with operators' knowledge, and interact with factory operators to support their problem-solving and decision-making. The CAs feature context awareness capabilities by integrating live data from the production line and operator activity tracking systems, as well as NLP technology (including LLMs). The dissertation offers insights into the opportunities, design, and implementation challenges, usability, user experience (UX), and socio-technical implications of deploying CAs in manufacturing settings. It also highlights the tensions between factory operators and management regarding engagement in knowledge sharing and digitization of the factories.

This dissertation contributes to open science to the extent that privacy and confidentiality agreements with our industrial partners and human participants allow[1].

[1] https://doi.org/10.4121/c2b34533-233b-4145-afd9-f2e0ad997736

SAMENVATTING

Moderne fabrieken vertonen weinig gelijkenis met de assemblagelijnen van een eeuw geleden. Tegenwoordig kan een enkele persoon verantwoordelijk zijn voor het opzetten, bedienen en onderhouden van een complexe keten van machines. Deze verschuiving heeft het werk van fabrieksoperators cognitief veeleisender gemaakt, wat constante monitoring en probleemoplossing vereist. Cruciaal voor het aanpakken van deze uitdagingen is de effectieve samenwerking van menselijke operators bij het uitwisselen van ideeën en kennis, van strategieën voor probleemoplossing tot oplossingen voor *emerging* problemen. Echter, de verhoogde productiviteitseisen en ploegdiensten betekenen minder mogelijkheden om deze kennis persoonlijk te delen en documentatiepraktijken krijgen minder prioriteit. Hierdoor kampt de maakindustrie met een crisis in *knowledge management*.

Dit proefschrift onderzoekt de integratie van *conversational AI assistants* in productieomgevingen om kennisdeling onder fabrieksoperators te faciliteren. Gebruikmakend van recente ontwikkelingen in Natural Language Processing (NLP), met name in *Large Language Models (LLMs)*, onderzoekt dit werk het ontwerpen en evalueren van *conversational AI assistants* die efficiënt menselijke kennis op de fabrieksvloer vastleggen en delen, terwijl ze tegemoetkomen aan de behoeften en zorgen van operators. De introductie van *conversational AI assistants* voor kennisdeling—hier aangeduid als *Cognitive Assistants (CA)*—in fabrieksomgevingen belooft aanzienlijke voordelen, maar brengt ook tal van uitdagingen met zich mee.

Door middel van een *research-through-design* voeren we etnografische studies, prototypeontwikkeling en empirisch onderzoek uit. Door het ontwerpen en inzetten van verschillende functionele prototypes in echte productieomgevingen, onderzoekt dit werk hoe *CAs* effectief kunnen worden ontworpen in overeenstemming met menselijke en organisatorische behoeften om kennisdeling in fabrieken te faciliteren. We hebben iteratief verschillende *CA* prototypes ontwikkeld om nuttige informatie uit (ongestructureerde) natuurlijke taal te extraheren, kennisbanken dynamisch bij te werken met de kennis van operators, en te communiceren met fabrieksoperators om hun probleemoplossing en besluitvorming te ondersteunen. De *CA's* beschikken over *context-awareness* door live data van de productielijn en *operator tracking* te integreren, evenals *NLP* technologie (inclusief *LLMs*). Het proefschrift biedt inzichten in de kansen, ontwerp- en implementatie-uitdagingen, bruikbaarheid, *User Experience (UX)* en *socio-technical* implicaties van het inzetten van *CA*'s in productieomgevingen. Het belicht ook de spanningen tussen fabrieksoperators en management met betrekking tot betrokkenheid bij kennisdeling en digitalisering van de fabrieken.

Dit proefschrift draagt bij aan *open science* voor zover privacy- en vertrouwelijkheid

met onze industriële partners en menselijke deelnemers dit toestaan[1].

[1] https://doi.org/10.4121/c2b34533-233b-4145-afd9-f2e0ad997736

1

Introduction

The era of factory workers performing mainly repetitive tasks is passing; they are now operators of complex systems. In contemporary manufacturing settings (see Figure 1.1), a single operator may be tasked with configuring, operating, and maintaining an intricate network of machines [1]. This transition necessitates continuous monitoring and problem-solving, markedly increasing the cognitive demands on factory operators [2, 3]. Effective collaboration among operators in sharing ideas and knowledge, from high-level problem-solving strategies to solutions to emerging issues, is key to tackling these challenges [4]. However, heightened productivity requirements and shorter shifts mean fewer opportunities to share this knowledge face-to-face [5]. Additionally, efforts to document knowledge are encouraged but not systematically executed [6]. Thus, effectively managing organizational knowledge in factories is a critical challenge.

Figure 1.1: A factory operator solving a problem with one of the machines along a detergent production line

An organization's success hinges on leveraging the knowledge found within the minds of its employees, customers, and suppliers—a process known as Knowledge Management (KM) [7, 8]. The benefits of sharing knowledge at scale are diverse: humans forget things, change jobs, and mentors are not always available. KM has become increasingly important in recent years due to the lack of skilled workers to replace retiring experts and the need for reskilling due to technological developments and the green transition in industries like manufacturing [9]. One key area of technological development is Artificial Intelligence (AI), which refers to computer systems that exhibit human-like intelligence in tasks such as reasoning or language understanding. Large Language Models (LLMs), trained on vast corpora of text data, have recently emerged as a prominent example of AI systems that have the potential to support diverse tasks from writing to decision-making [10]. In this work, we focus on (LLM-powered) Cognitive Assistants (CA), conversational AI systems that support cognitive tasks such as knowledge sharing and learning. Recognizing the labor crisis and the key role technology can play in mitigating it, the EU has recently announced a fund of €65 billion to invest in skills to take advantage of technological developments, such as ChatGPT, an LLM application (see Figure 1.2) [11].

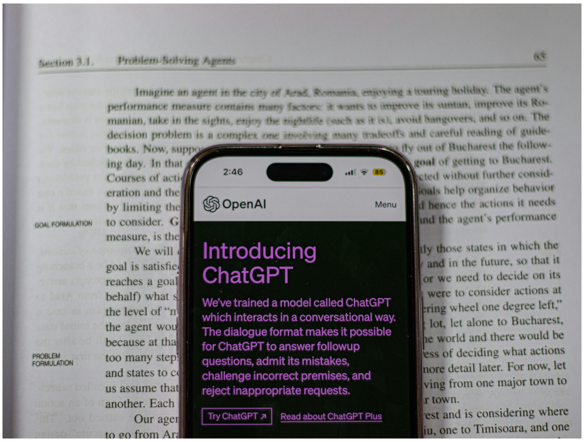

Figure 1.2: The release of ChatGPT in November 2022 marked a significant advancement in Natural Language Processing.

The importance of KM has motivated efforts to use AI technology, such as expert systems built on knowledge bases, to support human decision-making. This research field has over 60 years of history. Building expert systems involves a knowledge engineering cycle, where "knowledge engineers" apply knowledge elicitation techniques to human experts to formalize strategies and rules for the expert system [12]. However, the success of expert systems was often hindered by the challenge of creating and maintaining a high-quality knowledge base, known as the *knowledge acquisition bottleneck* [13]. Human knowledge can be complex and contextually linked, making it challenging to express and store. In dynamic environments, knowledge can quickly become obsolete [14]. Additionally, systems built to share knowledge face social challenges, such as the trade-off between perceived benefit and effort for the knowledge sharer, resulting in behaviors such as knowledge hiding [7]. Consequently, successfully implementing AI-driven Knowledge Sharing (KS) must tackle challenges related to human effort, the ephemeral and contextual nature of knowledge, and social barriers.

In recent years, Natural Language Processing (NLP) technologies, such as speech-to-text models and conversational AI, have been used effectively in high-demand contexts, such as generating transcriptions of medical interventions. However, using NLP to extract knowledge from human texts was challenging due to the variability of human language, such as inconsistencies in terms, pronoun usage, incomplete sentences, or missing context [15–17]. Recent advancements in NLP, particularly LLMs, have made it possible to effectively extract knowledge from unstructured texts, as demonstrated by their ability to answer queries about long, complex texts. The key innovation is the way LLMs consider the context of a word by paying attention to surrounding words [18]. As a result, NLP technologies can exhibit near-human-like abilities in conversation and reasoning, potentially playing a central role in tackling the challenges of knowledge acquisition for AI.

While effective implementation of conversational AI for knowledge acquisition is rare, there are many examples of conversational AI (i.e., chatbots) being used for information retrieval and decision support (e.g., [19–23]). Conversational AI applications provide an interactive means to find answers to problems, similar to how a human agent might help. In manufacturing, there are several examples of conversational AI being used, for example, to train new employees [19], facilitate easy information retrieval on the shop floor [20, 21], support diagnostics of machines [24], or deliver instructions for assembly tasks [22]. These examples demonstrate the potential of conversational AI to enable efficient and natural interactions with operators in the shop floor environment. They also highlight the *lack of knowledge on the real-world utility for factory operators, associated risks, and challenges with using conversational AI for knowledge sharing in manufacturing.*

Introducing new technologies to support operators in manufacturing is challenging due to several factors, such as the lack of unified IT systems [25], the high pace of work [1], and resistance from operators to change [6], especially if additional monitoring is involved [26]. Thus, it is important to consider these factors when developing new systems, such as CAs, so that appropriate designs and mitigation strategies can address them. For example, involving factory operators early in the development process can help design systems that accommodate the high pace of work and address the operator's needs. Thus, when designing and deploying AI systems in manufacturing, it is important to *consider the socio-technical challenges* and employ *human-centered design processes.*

Maintaining a socio-technical perspective, this work employs a *Research-through-Design (RtD) approach* to design and evaluate several iterations of CAs for factory operators. The main purpose of the CA is to facilitate efficient and effective knowledge sharing among factory operators to support their work operating, maintaining, and fixing the complex production line. Our overarching research aim is to answer the question: **How can we design an effective conversational AI knowledge sharing system for factory settings?**. This is further broken down into four sub-questions that represent the focus of the core chapters, namely:

RQ1 What are the opportunities and design challenges when deploying cognitive assistants to support knowledge sharing between factory operators? (Chapter 2)

RQ2 How do modality, user training, and context experience affect the user experience, usability, and interaction efficiency of cognitive assistants for knowledge sharing in factories? (Chapter 3)

RQ3 What are the implications of using large language model-based conversational AI for knowledge sharing among factory operators? (Chapter 4)

RQ4 What are factory operators' and management's perceptions of the impact and socio-technical risks and challenges of using cognitive assistants for knowledge sharing? (Chapter 5)

By pursuing an answer to these questions, this dissertation offers knowledge, methodological, and technical contributions, including design guidelines, working prototypes validated and evaluated in ecologically valid setups, and insights into the impacts, challenges, and risks of using CAs for factory operators. The remainder of this introduction goes into more detail regarding the context of this work, factory operations (1.1), the process of knowledge sharing (1.2), existing research on knowledge sharing systems (KSSs) (1.3), and underlying technologies for KSSs, such as LLMs (1.4).

1.1. MODERN FACTORY CONTEXT: OPERATORS, KNOWLEDGE NEEDS AND PROBLEMS

Modern manufacturing has been transformed by advanced digital technologies and automation, known as Industry 4.0 or Smart Manufacturing. This shift has increased the cognitive demands on factory operators and highlighted the need for effective Knowledge Sharing (KS) mechanisms. This section examines the evolving role of factory operators, the knowledge challenges they face, and the limitations of current KS practices. We will also explore how Cognitive Assistants (CA) can address these challenges, emphasizing the importance of a human-centered approach in their design and implementation. *By understanding these socio-technical dynamics, we aim to identify strategies that enhance KS in manufacturing.*

The Industry 4.0 paradigm is characterized by heightened connectivity, adjustability, and automation, aiming to streamline production processes to achieve cost reduction, enhanced quality, and better responsiveness to customer demands. A key feature of this era is the broad implementation of digital technologies to improve productivity and sustainability. Nevertheless, Industry 4.0 initiatives, while being primarily technology-focused, have often overlooked the critical component of human operators and their needs [27, 28]. Indeed, technologies should be aligned to fit human workers and enhance work satisfaction [26]. Addressing this gap, the emerging paradigm of human-centered manufacturing, or Industry 5.0, seeks to reconcile the capabilities of humans and machines, supporting factory workers in the increasingly knowledge-intensive manufacturing domain [28, 29].

In agile manufacturing, factories can rapidly switch between producing different products depending on external factors such as supply and demand. The complexity inherent in operating modern agile production lines—which comprise many machines operating in conjunction—poses a significant challenge (see Figure 1.4. Factory operators are no longer confined to monotonous manual tasks but are engaged in more cognitively demanding activities, such as dynamically adjusting machine settings, product-specific configurations, and solving problems with complex systems [1, 2]. These production line systems are susceptible to numerous potential failures and operate continuously amidst dynamic, noisy, and

Figure 1.3: Filled detergent canisters exiting the filling and capping machine in an agile factory.

high-stakes environments (see Figure 1.3). Consequently, even skilled operators find their work mentally taxing [30]. *Thus, providing operators with the knowledge and information relevant to the situation and their expertise is a key challenge [2].*

Despite the increasing use of digital technologies, the training and KS process in manufacturing remains heavily dependent on direct human interaction [31]. Traditional approaches—where novices learn from seasoned workers through extended one-on-one apprenticeships—though effective, are both time-consuming and costly. The reliance on such methods presents a challenge, particularly when experienced operators exit the workforce, taking with them invaluable knowledge [32], alongside risks such as forgetting or experts not being available when needed (see Figure 1.5). Coupled with the lack of skilled workers, this indicates a critical knowledge management crisis in manufacturing. In the US alone, the manufacturing sector will be short 2.1 million workers by 2030 due to skill gaps, which is expected to cost the US economy $1 trillion per year [33]. However, current practices to capture knowledge for training and KS purposes, including the attempt to manually codify knowledge [34] and store it in IT databases [35], are limited in scalability and sustainability because of how resource intensive the process is [5]. Furthermore, many factories lack the necessary IT infrastructure, have isolated data, and independent digital systems [36]. *In response to these challenges, there is a dire need for a centralized, efficient, and scalable KS in the manufacturing environment.*

While a knowledge management crisis might worry business owners or governments, these concerns are far removed from the day-to-day concerns of a factory operator. The operator's primary job is to ensure the production line runs effectively during their shift. For this, operators rely on their accumulated knowledge, rarely resorting to manuals, although they may ask an experienced colleague at times. Many factory operators are deeply proud of their expertise [31]. Besides the challenges of training new operators to replace retiring or leaving

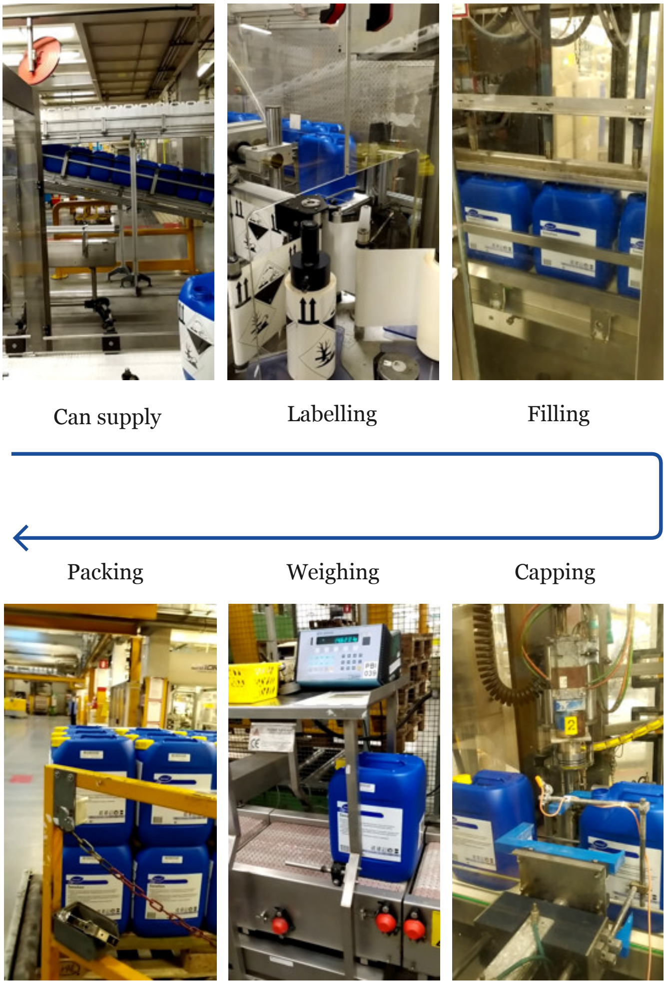

Figure 1.4: The main production steps on an agile production line for detergent cans. The line can operate 24/7 with one or two human operators.

personnel, factory managers also recognize the significant disparity between the production performance of existing operators. This disparity is unlikely to be attributed to a single factor but a combination of factors such as intelligence, skillfulness, motivation, training, or knowledge guarding. Emphasizing the social and organizational tensions in the workplace, factory operators sometimes purposely submit incomplete or inaccurate production data to ensure management does not have an accurate representation of their performance [37]. Additionally, some operators are intent on guarding their knowledge to maintain a position of power [31]. These challenges underscore the importance of considering the social and organizational context while designing a new tool, such as a CA. *Therefore,*

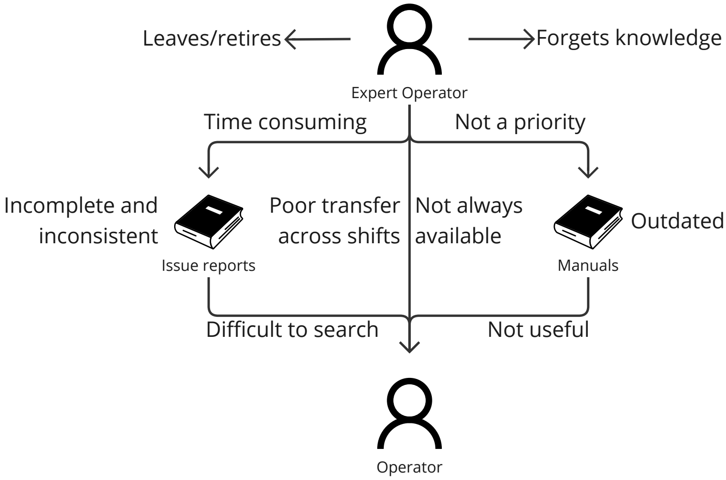

Figure 1.5: The current knowledge flow in factories (depicted by arrows) and the social and practical challenges for effective knowledge sharing.

system designers should consider the needs of the operators on the shop floor, both practically for their work and socially as members of a team and larger organization.

In summary, traditionally, manufacturing has focused on automation to increase productivity without considering the valuable contribution of human operators or their needs [38]. The transition to a more human-centered perspective presents various challenges and opportunities for Knowledge Management (KM) [39]. *As operations become increasingly knowledge-intensive and the lack of skilled workers more severe, the need for effective, human-centric, and scalable knowledge-sharing mechanisms becomes more critical.* The subsequent discussion on KM strategies, technological innovations, and socio-technical systems aims to identify solutions to these challenges, leveraging the combined strengths of human intellect and technological advancements.

1.2. KNOWLEDGE SHARING: STRATEGIES, TECHNOLOGY AND SOCIAL FACTORS

Knowledge management (KM) is an interdisciplinary field that integrates concepts from cognitive psychology, information sciences, educational psychology, business, and manufacturing. This entails processes to acquire, manage, and share knowledge within an organization to achieve a competitive advantage [8]. In practice, the core tenet of KM is to *record things that might be useful later* [40]. The application of KM in organizations can lead to significant improvements in efficiency, innovation, and problem-solving capabilities [7]. *However, designing an effective KM system to realize these benefits, particularly in dynamic environments such as factories involving complete socio-technical systems [41], operators may resist new digital tools [37], and* *documentation practices are often undervalued [16, 31, 42], remains a substantial challenge.*

In the following sections, we will outline the key processes surrounding Knowledge Sharing (KS) when facilitated by a digital system such as a CA, namely, knowledge acquisition, sharing, and application.

Knowledge acquisition involves eliciting knowledge, explicating it, and formalizing it for later use [43]. Knowledge can be elicited once it has been created in the mind of an employee. *Eliciting* involves extracting knowledge from domain experts or sources, often through interviews, surveys, observation, or analysis of existing documentation. Elicitation can uncover both explicit and tacit knowledge, understanding not just what is known but also how it is applied in practice. In this work, we focus on explicit knowledge that can be readily expressed in words as opposed to tacit knowledge that can not [44].

Once knowledge has been elicited, it needs to be made explicit. *Explication* involves articulating the knowledge clearly and understandably. This may involve defining terms and organizing information into coherent structures. Explication helps ensure that knowledge is comprehensible to others and can be effectively communicated.

Finally, during *formalization*, knowledge is structured and represented in a formal language or framework. Formalization makes knowledge more precise and facilitates analysis and manipulation by digital systems. This might involve creating models, rules, ontologies, or other formal knowledge representations. Formalization helps facilitate (AI) reasoning and decision-making based on the captured knowledge.

Knowledge sharing is the process of transferring knowledge between individuals, groups, or organizations [45]. Traditionally, this can be done through various channels such as documentation, training sessions, workshops, seminars, meetings, or digital platforms. While many of these strategies aim to stimulate face-to-face KS, organizations are also interested in systems that formalize the knowledge. This ensures that the knowledge persists beyond employees' memory and facilitates sharing at scale and asynchronously. In this work, we explore using CAs as an intermediary to help efficiently share newly created knowledge.

Knowledge application is when knowledge is used to solve problems and make informed decisions to profit the organization [46]. This involves analyzing the situation, identifying relevant knowledge, and applying it effectively to address the issue. In the context of this work, the CAs contribute primarily to identifying and retrieving relevant knowledge. Analyzing the situation and taking action is still largely up to the human operator.

1.3. Designing Knowledge Sharing Systems

Designing KS systems requires consideration of a broad range of factors, often intertwined with social and human elements. Organizations across various sectors have developed their own dedicated KS systems. For instance, the 'Inside IBM' project used AI, information systems, and user-centered design to aggregate IBM's accumulated product support knowledge into a single system [47]. Siemens employs TechnoWeb for social collaboration, which leverages a combination of social networking tools, collaborative platforms, and knowledge repositories to facilitate the sharing and management of expertise across the organization [48]. TechnoWeb integrates features such as discussion forums, expert directories, and document sharing to create a dynamic and interactive environment for knowledge exchange. NASA uses the Lessons Learned Information System (LLIS)[1] to document and share project experiences. LLIS employs a structured database approach, where lessons are systematically categorized, indexed, and made searchable to ensure that critical insights from past projects are easily accessible for future missions [49]. This system includes detailed metadata tagging, cross-referencing of related lessons, and a review process to validate the accuracy and relevance of the information. Furthermore, the enhanced version of LLIS that the Goddard Space Flight Center uses employs an 'active' push feature to recommend lessons that match the knowledge needs of specific individuals. *These diverse systems illustrate the various methods employed to manage and leverage knowledge within organizations, ranging from searchable lessons learned and social collaboration tools to active knowledge recommendations and systematic documentation processes.*

Several factors contribute to the success of KS systems, as exemplified by the cases of Google, Toyota, and Xerox's Eureka project described below. In the case of Google, they emphasize psychological safety for KS [50]. Toyota—perhaps KM's most renowned manufacturing success story—employs network-level knowledge-sharing processes. These processes effectively engaged members in sharing critical knowledge, deterring free-riding behavior, and minimizing the obstacles to locating and obtaining valuable knowledge [51]. The Eureka project at Xerox showcases the successful use of technology to gather, validate, and share best practices across their customer service engineers [52]. By involving customer service engineers throughout the design process and encouraging local champions to support adoption and feedback, Xerox contributed to the project's success [53]. Ultimately, the Eureka project is said to have saved Xerox $100 million [52, 53]. *Overall, these cases highlight the importance of minimizing the burden on knowledge contributors, ensuring high knowledge quality, and involving end-users throughout the design process to address their needs.*

KS systems incorporating AI and Internet of Things (IoT) technologies are transforming KS within manufacturing environments [19]. Yet, despite the promise of these technologies to enhance KS systems, several socio-technical hurdles persist. For instance, significant resources are required to develop and maintain knowledge bases [54], and data quality issues often undermine efforts to automatically uncover knowledge from existing unstructured issue reports [15–17]. Similar challenges have faced the use of NLP in healthcare, such as IBM's Oncology Expert Advisor in 2016, which searched through 'unstructured' physician notes to help suggest treatments. Yet, Watson's NLP failed to effectively process the jargon, shorthand, and subjective comments or pick up the nuances that human physicians could, demonstrating the challenges in handling human-generated texts [55]. These challenges are socio-technical because they involve both technical aspects, such as data quality

[1] https://llis.nasa.gov/—last accessed November 20, 2024

and the capabilities of NLP in processing it, and human aspects, such as the effort put toward documentation, which inhibits the data's utility for KS [42]. *Tackling the human aspect, there is a knowledge gap in designing KS systems that facilitate the acquisition of human knowledge while balancing the (perceived) effort of knowledge authoring such that factory operators are consistently engaged in contributing useful knowledge.*

1.4. Conversational AI for Knowledge Sharing among Factory Operators

Knowledge sharing systems must support efficient and reliable interactions to provide utility to factory operators while minimizing the perceived effort involved. In recent years, conversational interfaces have enabled efficient communication between users and digital systems in natural language, fundamentally transforming how humans interact with technology. Unlike traditional graphical user interfaces that require interaction through visual elements like buttons and menus, conversational interfaces allow users to engage in dialogue using text or voice. These interfaces can understand and respond to queries, execute commands, and provide personalized assistance. This mode of interaction leverages Natural Language Processing (NLP) techniques, making interactions more flexible and accessible. These interfaces represent a key development in human-computer interaction, especially when powered by state-of-the-art NLP, such as Large Language Models (LLMs).

Research into AI assistants in manufacturing has recently increased, taking advantage of recent advancements in NLP and the proliferation of IoT to support decision-making, operationally efficiency, and training [19, 27, 56–61]). In the following sections, we discuss research on AI assistants in manufacturing, the use of knowledge bases, knowledge acquisition from humans for KS, and the associated socio-technical challenges.

1.4.1. AI Assistants in Factories

AI assistants are digital tools designed to support humans by completing tasks or providing information. For example, IBM Watson, which became famous in 2011 when it won Jeopardy!—navigating a complicated game full of wordplay to search a textual database for the correct answer and respond naturally [55]. In factories, AI assistants working with an AI assistant form a socio-technical system consisting of the operator, the operator's tasks in the factory, and the technology enabling the assistant [62]. AI assistants can help operators fix problems with the machines or set them up optimally by enabling efficient information retrieval, remote machine control, decision-support, and sharing knowledge [60, 63–65]. To enhance their capabilities, AI assistants can be integrated into factory systems, such as control and monitoring systems for production machines, scheduling systems, and sensors. Some systems make use of the available data to perform predictive maintenance using ML [66] or NLP to extract and represent knowledge from texts [67]. *To interact with operators, most AI assistants use conversational AI—a chat or voice interface—to* *provide efficient and natural ways for operators to communicate.*

1.4.2. KNOWLEDGE-INTENSIVE AI ASSISTANTS

Knowledge-intensive AI assistants are a subset of AI assistants characterized by their ability to utilize extensive knowledge bases and perform complex reasoning tasks. The knowledge bases and rules that inform these assistants in manufacturing are usually static, having been defined by a domain expert during development [19, 21, 68, 69]. Some systems use live data but are focused on presenting the status of production systems and perform (limited) reasoning over the data [20, 70–73]. For example, Casillo *et al.* [19] presents a chatbot for training new operators based on a predefined curriculum, obtaining promising results. Baldauf *et al.* [73] deployed smartwatches that notified operators of machine errors, citing the importance of concise information display. Additionally, the chatbot demonstrated by Trappey *et al.* [21] utilizes VR to deliver knowledge in response to Frequently Asked Questions (FAQs). *Overall, existing research on AI assistants focused on understanding how to deliver knowledge or present (live) information to operators but not acquire knowledge from operators.*

1.4.3. COGNITIVE ASSISTANTS: ACQUIRING AND SHARING HUMAN KNOWLEDGE

In the context of this dissertation, **Cognitive Assistants (CAs) are an advanced type of knowledge-intensive AI assistant that interacts conversationally to acquire knowledge from humans, share it with other humans, and support knowledge application.** Cognition refers to the mental processes of acquiring and comprehending knowledge [74], which the CA aims to support. To do so effectively, a CA relies on several technologies, including dialogue management, NLP, context awareness, databases, and ontologies. To maximize utility for factory operators, CAs can be integrated with factory systems and operator workflows, making them a cyber-physical system [31] (see Figure 1.6).

Acquiring knowledge using conversational AI and/or mobile devices has been explored and shown to be promising. For example, Fenoglio *et al.* [75] introduced a role-playing game involving virtual agents, human experts, and knowledge engineers to refine knowledge graphs that were algorithmically generated. The authors raised several important ethical concerns regarding the processing of personally identifying data, deciding to avoid audio and video recordings, and other concerns regarding misuse of employee monitoring. Hoerner *et al.* [58] built a digital assistance system to support operator troubleshooting processes on the shop floor using captured knowledge. While they allowed operators to suggest edits, *knowledge engineers still performed the initial knowledge capture process.* In contrast, Hannola *et al.* [27] deployed smartphone applications for factory operators to take notes and videos of solutions to production line problems themselves, thus enabling other operators to access newly created knowledge. *While this work highlighted the potential of facilitating KS through digital technologies, it did not investigate the application of LLMs and/or conversational AI.*

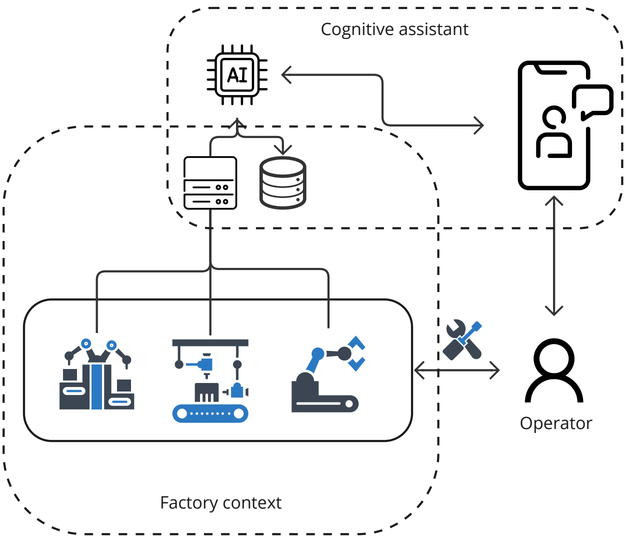

Figure 1.6: The scope of this dissertation includes the factory context, operator and cognitive assistant system

Examples outside of manufacturing demonstrated the successful use of conversational AI to crowdsource knowledge acquisition, namely, a game to elicit knowledge from crowdworkers [76] and a context aware chatbot that simultaneously fulfills information needs while acquiring new knowledge [77]. A key consideration for these systems is ensuring the quality of acquired knowledge, thus employing rating or validation mechanisms. *Although these conversational AI systems were not deployed in a manufacturing environment, they demonstrate the potential of using conversational interactions to efficiently acquire knowledge.*

1.4.4. Socio-Technical Challenges for Deploying Cognitive Assistants in Factories

Work in factories is influenced by social aspects and organizational relations rather than purely technical, deterministic activities [78], making introducing new tools or workflow adjustments challenging [41]. Digital tools can facilitate KS among operators; for instance, enabling operators to post comments under existing work instructions [79]. However, these tools can have many unforeseen effects, face social barriers, and raise ethical implications. For example, even introducing relatively simple tools, such as a repair ticketing system, can shift the balance of control towards or away from operators depending on who is given privileges to create, view, and modify tickets [78]. Tools for decision support can also have adverse effects if not designed appropriately with human cognitive limits in mind, such as information overload [5]. While decision support tools can benefit inexperienced operators by providing guidance and allowing independent operation, they may be redundant for experienced operators relying on their expertise (see Figure 1.7) [5].

Figure 1.7: A novice operator observes how an experienced operator solves a problem after requesting help.

Furthermore, it is important to consider the context factors surrounding the knowledge being shared [8], which is often overlooked when deploying (agential) AI-powered systems [80]. These context factors include the organizational culture, the psychological readiness of the workforce, and the existing social dynamics within the team [31]. For instance, a system that does not account for the hierarchical structure of an organization may inadvertently reinforce existing power imbalances, leading to resistance from employees [37]. Similarly, neglecting the psychological readiness of factory operators—such as feelings of losing control [2], trust in the system, perceived usefulness [81], and levels of anxiety or resistance [82]—can result in mistrust towards the new technology, thereby hindering its adoption.

Despite the recognition of these factors, there remains a significant gap in the literature regarding the specific socio-technical elements that influence the successful adoption of conversational AI KS systems. Specifically, *there is a lack of real-world studies examining adoption challenges for these technologies from the perspective of factory operators.* Addressing this gap is essential for developing systems that are not only technologically advanced but also socially and psychologically aligned to the needs of factory operators.

1.4.5. Natural Language Processing and Other Technologies that Power Cognitive Assistants

While previous attempts to use AI for KS, namely expert systems, were held back by the lack of context awareness, rigid knowledge representation, and one-way knowledge transfer from expert to novice, newer NLP technologies and context awareness can tackle these shortcomings [31]. Modern NLP technologies can provide a more nuanced understanding of a factory operator's knowledge, enhanced by additional data streams such as machine error codes or human tracking. Thus enabling efficient knowledge acquisition at the production line. Modern NLP techniques for conversational AI fall under two paradigms, traditional intent- and LLM-based systems, exhibiting trade-offs in adaptability and reliability, among other aspects.

Traditional Conversational AI: Intent-based Systems

Underpinning many modern conversational AI assistants are intent-based techniques grounded in the principles of symbolic AI and rule-based processing. These systems are designed to recognize and interpret user intents through predefined patterns and commands [83, 84]. They rely heavily on structured knowledge bases and decision trees, enabling them to respond to specific queries or execute tasks based on recognized commands. While limited in flexibility and adaptability, this approach provides a high degree of control and predictability in interactions, making it suitable for applications where precision and rule adherence are important. However, the rigidity of intent-based systems poses significant challenges, especially in understanding nuanced phrasing [85]. Additionally, their rigidity can cause problems in dynamic environments where human needs and contexts rapidly evolve. Furthermore, they are time-intensive to build because the system's behavior is largely explicitly defined. *In short, intent-based systems offer control and predictability at the costs of intuitiveness, flexibility, and development and maintenance efforts.*

State-of-the-Art Conversational AI: Large Language Model-based Systems

In contrast to the rigid intent-based systems, large language models (LLM)-based systems can be highly flexible. LLMs such as Llama3[2], Claude 3.5 Sonnet[3], or GPT-4-omni[4](see Figure 1.8) can understand, generate, and interact with human language at a highly sophisticated level. Built upon the transformer architecture, which employs self-attention mechanisms, LLMs mark a significant step forward from their predecessors, such as Recurrent Neural Networks (RNNs) [86]. Unlike RNNs that process text sequentially and often struggle with long-range dependencies, LLMs can analyze and generate text in parallel, allowing them to handle extensive context and complex language patterns efficiently [87–89]. This makes them more

[2] https://llama.meta.com/llama3/—last accessed November 20, 2024
[3] https://www.anthropic.com/news/claude-3-5-sonnet—last accessed November 20, 2024
[4] https://openai.com/index/hello-gpt-4o/—last accessed November 20, 2024

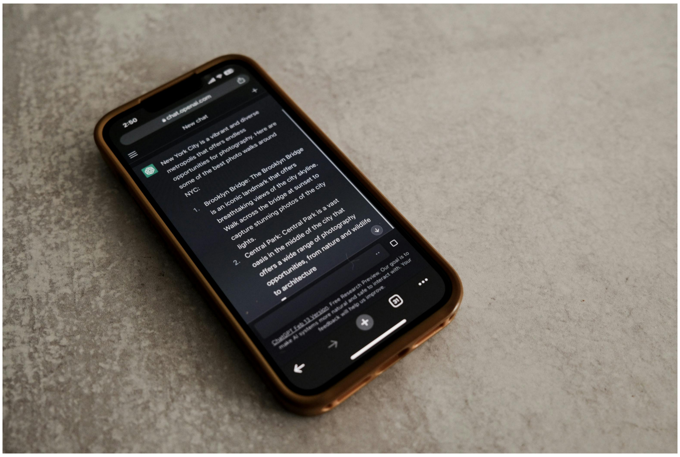

Figure 1.8: ChatGPT's chat interface, the most prolific LLM application at the time of writing.

powerful and versatile than intent-based systems, which are typically rule-based or utilize simpler machine learning models focused on understanding specific user intentions in limited contexts.

LLMs surpass the capabilities of intent-based systems in terms of adaptability and depth of understanding, enabling more reliable and nuanced conversational interactions [90]. These qualities are especially beneficial when using AI-based systems, such as CAs, to support problem-solving and decision support for complex systems, such as manufacturing production lines. The flip side of an LLM's adaptability is that its output is non-deterministic and cannot be fully trusted to be accurate, for example, "hallucinated" responses that appear plausible but are not faithful to the provided material [91]. In factories, hallucinated responses could severely impact production performance, quality, and/or operator safety. Despite these concerns, the potential benefits of LLM-based systems for supporting KS in manufacturing are evident. The knowledge for operating complex agile production lines is ephemeral and context-specific, and there is an abundance of unstructured text-based documentation. Yet, these predictions are yet to be comprehensively confirmed in the literature, *necessitating system development and thorough evaluation in the real-world investigating their utility for factory operators and associated risks and challenges.*

1.4.6. LARGE LANGUAGE MODEL-POWERED SYSTEMS IN MANUFACTURING

Taking advantage of the advanced NLP capabilities of LLMs, factory operators may benefit by harnessing the knowledge in their extensive corpus of text documentation, such as work instructions or machine manuals, complemented by the ongoing addition of knowledge shared by operators. Although unstructured issue reports were previously considered challenging for natural language processing due to issues such as poorly structured text or inconsistent terminology [15, 16], advancements in NLP such as LLMs can now effectively address these challenges.

Manufacturers are cautiously adopting Large Language Models (LLMs) while addressing associated risks. For instance, Mercedes-Benz [92] are experimenting with ChatGPT in vehicle production, enhancing error identification, quality management, and process optimization. This AI-driven method allows quality engineers to simplify complex evaluations and presentations of production data through dialogue-based queries. Researchers are also investigating the use of LLM-powered tools in knowledge-intensive manufacturing scenarios. Xia *et al.* [93] showcased how in-context learning and the injection of task-specific knowledge into LLMs can improve the planning and control of production processes. A systematic test by Wang *et al.* [94] evaluated ChatGPT's responses to 100 manufacturing-related questions. They assessed the responses for correctness, relevance, clarity, and comparability, highlighting areas for improvement such as low scores in critical analysis, occasional non-factual responses, and dependency on query quality, demonstrating the shortcomings of using foundational models for specific knowledge domains.

While foundational LLMs do not contain context-specific knowledge—for example, how to fix a machine in a specific factory—they possess extensive general knowledge and reasoning abilities [88], enabling them to excel in processing complex information [87], generate insights and reasoning [89]. Nevertheless, LLMs face two key challenges in context-specific applications: (1) reliance on outdated information from their pre-training data, and (2) potential inaccuracies in factual content, a phenomenon termed "hallucination" [88, 95]. To mitigate these issues and maximize the utility of LLMs in specialized, knowledge-intensive tasks, techniques such as fine-tuning, chain-of-thought [96] few-shot prompting [97, 98], and Retrieval Augmented Generation (RAG) [99] can be adopted. While all these techniques can be used in conjunction to maximize utility, RAG is the only way to reliably add new knowledge to an LLM-based system without training a new model from scratch—an extremely resource-intensive process.

Instead of training a new model from scratch, RAG represents a low-cost, transparent, and reliable mechanism that supports operators with information when they request help with a problem. The RAG pipeline starts with retrieving relevant information from a comprehensive corpus, including documents such as technical manuals, issue reports, or knowledge graphs. Following retrieval, this information is automatically integrated into the prompt for the LLM. Through this integration, the LLM generates responses based on specific, contextual information from the retrieved documents. Factory operators can trace the source of the information used by the LLM, as the response is directly linked to the documents from which it was generated. This allows for easier verification, enabling humans to cross-reference the LLM's response with the original documents to ensure accuracy and relevance. In this way, RAG transforms the LLM into a more reliable tool, adapting its responses to include specific, contextually relevant information.

Thus, a CA with RAG-based decision-support combined with efficient (conversational) knowledge acquisition processes is a promising direction for enhancing KS in manufacturing. *However, there is a need to evaluate RAG-based CAs in the real-world to evaluate their utility for factory operators.*

1.4.7. RESEARCH OBJECTIVE

In summary, modern factories face significant challenges in KM despite advancements in digital technologies and automation. Heightened productivity requirements and fewer face-to-face interactions limit KS, and efforts to document knowledge are inconsistent. Expert systems struggle with the knowledge acquisition bottleneck, and human knowledge is complex and contextually linked, making it difficult to store and quickly obsolete. NLP technologies, such as LLMs, show promise but face challenges due to language variability and incomplete context in manufacturing texts. Additionally, the adoption of new technologies is hindered by factors such as a lack of unified IT systems, high work pace, and operator resistance, highlighting a significant gap in understanding the socio-technical elements crucial for successfully implementing conversational AI knowledge-sharing systems. Among this plethora of relevant and concrete challenges, this dissertation addresses the following knowledge gaps:

- Designing cognitive assistant systems for knowledge sharing: Investigating how to design systems that enable the effective sharing of factory operator knowledge within acceptable levels of effort for the operators (see Section 1.3).
- Evaluating LLM-based systems for knowledge sharing in manufacturing: Assessing the need for real-world evaluations of LLM-based systems to determine their utility for factory operators and to investigate associated risks and challenges (see Section 1.4.6).
- Understanding socio-technical elements: Exploring the specific socio-technical elements that influence the successful adoption of cognitive assistant for knowledge sharing in manufacturing environments (see Section 1.4.4).

By addressing these gaps, this dissertation contributes to the fields of Human-Computer Interaction (HCI), Computer-Supported Cooperative Work (CSCW), and Knowledge Management (KM) by providing design guidelines and insights into the impacts, challenges, and risks of using CAs for factory operators.

1.5. RESEARCH QUESTIONS AND APPROACH

Previous work has demonstrated numerous, often insurmountable, challenges when successfully deploying AI-powered knowledge management systems in manufacturing and elsewhere, such as the high authoring burden on experts and shortcomings of traditional NLP technologies in handling unstructured texts. Informing the design of a system to address this problem, we define the following overarching research question: **How can we design an effective conversational AI knowledge sharing system for factory settings?**

To ground this research in real-world experiences and realistic technological capabilities, we conduct most of the research in factories using functional prototypes, enabling the collection of ecologically valid user evaluations. Considering that work in factories is socio-technical rather than purely technical and deterministic [58, 78], we conduct our research in an exploratory, human-centered, and design-driven manner, as discussed below.

1.5.1. Approach: Research-through-Design

We adopt Research-through-Design (RtD) as our overarching methodological framework. RtD is a practical approach that emphasizes iterative design and prototyping. RtD allows us to engage with factory operators to learn more about their needs and the factory operations context, foster creative solutions, and keep track of design decisions and feedback. After creating the initial concepts informed by ethnographic and contextual studies, we construct prototypes at increasing levels of fidelity, culminating in fully functional CA systems integrated with the production environment. Prototypes serve as interactive probes, revealing insights about user interactions, system performance, and the socio-technical dynamics within the manufacturing environment. These prototypes are tested in the field, directly within the factory setting, allowing us to observe how factory operators interact with the system, how it integrates into their workflow, and how effectively it supports their work. Furthermore, testing in the field helps identify any unforeseen technical, organizational, or social issues, providing a comprehensive understanding of the system's impact on the factory floor and challenges. This method aligns well with our goal of learning what constitutes a successful CA that is attuned to the needs, behaviors, and contexts of factory operators.

To inform the RtD process, we apply guidelines and tools from various multidisciplinary fields, namely a combination of Socio-technical Systems Design (STSD), Human-Centered Artificial Intelligence (HCAI), Human-Computer Interaction (HCI), and Computer-Support Collaborative Work (CSCW).

STSD is an approach that emphasizes the integrated design of both the social system (including the organizational structure, people, and processes) and the technical system (comprising technology, machinery, and tools) within work environments [3]. The objective of STSD is to enhance overall organizational performance and worker satisfaction, although these may be at odds at times. This approach recognizes that neither the technical nor the social system operates in isolation; rather, they are interdependent and must be designed jointly to support each other.

CSCW (Computer-Supported Cooperative Work) extends this perspective by focusing on how collaborative activities and their coordination can be supported by computer systems [100]. It examines the ways in which technology can facilitate teamwork, communication, and collaboration among individuals and groups within an organization. CSCW emphasizes the importance of designing systems that support social interactions and collaborative processes, ensuring that technology enhances rather than hinders cooperative work. By integrating CSCW principles, STSD can better address the complexities of modern work environments where collaboration and communication are critical.

HCAI, on the other hand, supports the design of AI technologies that are designed with human welfare as a central consideration [101]. It is about creating AI systems that augment, rather than replace, human capabilities, ensuring that these systems are transparent, explainable, and trustworthy. HCAI stresses the importance of AI systems being adaptable to human needs and working effectively as part of human-machine teams.

HCI is a broader field that studies how people interact with computers and designs technologies that fit human contexts. It covers a wide range of topics, from usability and user experience design to the social implications of technology. HCI is primarily concerned with making technology accessible, usable, and beneficial to humans.

When addressing the challenge of designing CAs for factory settings, integrating principles from STSD, HCAI, HCI, and CSCW can provide valuable perspectives and approaches. STSD offers insights into the interplay of human, social, and organizational factors, which can inform the design process. HCAI emphasizes the importance of creating conversational AI technologies that augment human abilities, aiming to develop systems that act as reliable, trustworthy assistants rather than replacements for human expertise. HCI principles guide the design process to create user-friendly, engaging systems that effectively fit the operators' work context. CSCW adds an additional layer by focusing on the collaborative aspects of work, ensuring that the designed systems support teamwork, communication, and coordination among workers. While these methodologies do not guarantee specific outcomes, they offer frameworks and perspectives to support the development of well-rounded, human-centered solutions that enhance both individual and collective performance in factory settings.

1.5.2. PROJECT CONTEXT: FACILITATING FIELD RESEARCH AND INTERNATIONAL COLLABORATIONS

This work occurred in the context of the COALA project "COgnitive Assisted agile manufacturing for a LAbor force supported by trustworthy Artificial Intelligence", a European Commission Horizon 2020 research and innovation program project, which allowed us to conduct extensive fieldwork at factories. We conducted most of our field research at the four factories and vocational schools associated with the COALA project. Our primary factory was in Enschede, The Netherlands, and the rest near Milan, Italy. The access to working operators at factories enabled us to conduct research with high ecological validity. Furthermore, the project facilitated collaborations with our partners at BIBA (Bremer Institut für Produktion und Logistik)[5] on an open-source framework for speech-enabled conversational assistants; partners at ANITI (Artificial and Natural Intelligence Toulouse Institute)[6] on generating explanations for machine learning models; and partners at the University of Bremen[7] for developing didactic concepts for training factory workers.

[5] https://www.biba.uni-bremen.de/—last accessed November 20, 2024
[6] https://aniti.univ-toulouse.fr/—last accessed November 20, 2024
[7] https://www.uni-bremen.de/

Overall, the access and collaborations possible through the COALA project supported us in rigorously exploring the socio-technical perspective in the field and using state-of-the-art technologies, improving our work's ecological validity and societal relevance.

1.5.3. Research Questions and Chapter Overview

This dissertation comprises four central chapters (2-5), one for each research question that explores a different aspect of the above-mentioned overarching research question. The chapters cover the opportunities, design challenges, design optimization, implications, and perceptions of CAs for factory operators (see Figure 1.9). The last of these central chapters, Chapter 5, is a collection of longitudinal insights, representing a comprehensive guide for future work on CAs for KS in manufacturing. Chapter 6 rounds off this dissertation with a discussion of the answers to the research questions, implications for practitioners and researchers, reflections on the approach, and suggestions for future work.

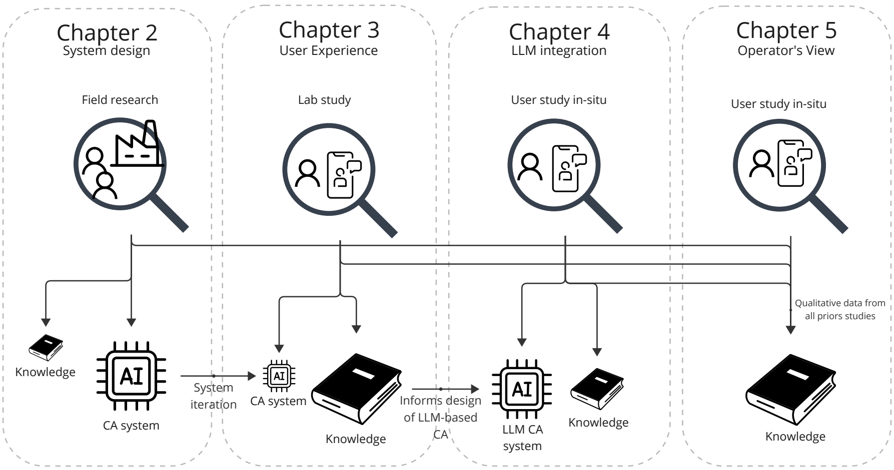

Figure 1.9: Overview of how the chapters are positioned and how the knowledge and technical contributions inform the work in subsequent chapters. For example, the system evaluated in Chapter 3 is based largely on the design proposed in Chapter 2, and the knowledge created from Chapter 2 and Chapter 3 is used to inform the system's design in Chapter 4. Lastly, Chapter 5 is a longitudinal perspective of the evaluations of all prototypes, culminating in comprehensive design guidelines.

Chapter 2: System Design and the Factory Context

The application of CAs in factories is a critical evolution of knowledge management systems made possible by recent advancements in NLP. Despite the recognized potential, the design and deployment of these systems in complex environments like factories present unique challenges. This observation leads us to our first research question. **RQ1: What are the opportunities and design challenges when deploying CAs to support knowledge sharing between factory operators?** To answer this question, we conducted ethnographic studies at several factories, leading to a CA design, usage scenarios, and identified challenges. Through several rounds of co-design sessions with factory representatives, we designed a CA system to provide human-centered knowledge management for factory workers. The proposed system employs conversational interaction to collect information regarding issue resolutions such that it can be represented in a knowledge graph; for example, event A causes event B, and event A can be solved by performing action C.

As such, the resulting system design is the primary contribution. Additionally, we contribute knowledge regarding promising usage scenarios for CAs during factory operations and the human, technical, and ethical challenges involved. This chapter is based on the following article: [102] *S. Kernan Freire, S. S. Panicker, S. Ruiz-Arenas, Z. Rusák, and E. Niforatos. "A Cognitive Assistant for Operators: AI-Powered Knowledge Sharing on Complex Systems". In: IEEE Pervasive Computing (2022), pp. 1–9. DOI: 10.1109/MPRV.2022.3218600.*

CHAPTER 3: THE EFFECTS OF DESIGN AND EVALUATION DECISIONS ON USER EXPERIENCE

Having identified the fundamental opportunities and design challenges in Chapter 2, we now turn to building a better understanding of the impact of design choices and tool introduction on user experience, leading to the second research question: **RQ2: How do modality, user training, and context experience affect the UX, usability, and interaction efficiency of CAs for KS in factories?** We developed a functional CA and accompanying machine interfaces to conduct a controlled empirical study. The participants included factory personnel, students, and HCI researchers. The scientific contributions include new insights into the effects of modality, training, and context expertise on interaction efficiency, UX, and usability. The most notably significant effect we measured was that participants from the factory context were significantly more positive regarding the usability and UX of the system compared to laymen. Furthermore, we qualitatively analyzed user feedback to identify areas for improvement. The analysis revealed a strong emphasis on enabling efficient and reliable interactions between the CA and the factory operators.

Additionally, the chapter builds on the technical contribution from Chapter 2, incorporating context awareness from machine data provided by a simulated production line. Lastly, we contribute knowledge through lessons learned and design guidelines informed by the qualitative feedback collected from the user study participants. The contents of this chapter are based on the following article: [103] *S. Kernan Freire, E. Niforatos, C. Wang, S. Ruiz-Arenas, M. Foosherian, S. Wellsandt, and A. Bozzon. "Lessons Learned from Designing and Evaluating CLAICA: A Continuously Learning AI Cognitive Assistant". In: Proceedings of the 28th International Conference on Intelligent User Interfaces. IUI '23. Sydney, NSW, Australia: Association for Computing Machinery, 2023, pp. 553–568. DOI: 10.1145/3581641.3584042.*

Chapter 4: Harnessing State-of-the-Art Natural Language Processing Techniques

Building upon the insights gained from optimizing the design of CAs, our attention shifts towards exploring the capacity of Large Language Models (LLMs) to further improve their usefulness. Chapter 4 introduces a CA system leveraging LLMs. The system employs Retrieval Augmented Generation (RAG) to efficiently respond to operator queries and supports knowledge authoring through a dynamic issue reporting form using the 5-whys method [104], coupled with LLM-powered knowledge validation. A field study in a detergent factory demonstrated the system's usability and utility.

The chapter addresses the research question: **RQ3: What are the implications of using LLM-based conversational AI for knowledge sharing among factory operators?** It explores the effectiveness of LLM-powered systems in extracting knowledge from poorly structured documents, noting that some operators still prefer human assistance when available. The chapter makes a system contribution of an LLM-based CA for factory operators. Through a benchmarking study, we compare various proprietary and open-source LLMs in handling manufacturing-specific queries. This benchmarking highlights the narrowing performance gap between open-source and proprietary models, offering insights into selecting suitable LLM tools for knowledge management in manufacturing, considering performance, privacy, and customization needs.

Overall, the chapter contributes to an understanding of the impact, opportunities, and risks of LLM-based CAs for factory operators and provides a system design for LLM-powered KS and interaction. The chapter is based on the following article: [105] *S. Kernan Freire, C. Wang, M. Foosherian, S. Wellsandt, S. Ruiz-Arenas, and E. Niforatos. "Knowledge sharing in manufacturing using LLM-powered tools: user study and model benchmarking". In: Frontiers in Artificial Intelligence 7 (2024). ISSN: 2624-8212. DOI: 10.3389/frai.2024.1293084.*

Chapter 5: Contrasting Perspectives of Operators and Managers Regarding CAs for Factory Operators

While technological advancements can potentially transform knowledge-sharing practices, the successful integration of CAs within factory settings hinges on a deep understanding of the social and organizational context. With this in mind, we formulate the final research question: **RQ4: What are factory operators' and management perceptions of the impact and socio-technical risks and challenges of using cognitive assistants for knowledge sharing?** To investigate these issues, we compiled comments and observations of factory operators and managers from over two years of designing, deployments, and evaluations in two detergent factories. In turn, we conducted a hybrid deductive and inductive thematic analysis [106, 107]. The analysis confirmed the need for systems like CAs and the potential benefits and apprehension regarding safeguarding knowledge quality, privacy issues, and user adoption.

As such, this chapter presents knowledge contributions regarding the differing perceptions of operators and their managers toward (LLM-powered) CAs, including the expected impacts, risks, and design considerations. These insights, which stem from a longitudinal study, are translated into design guidelines to facilitate future development and research work. The contents of this chapter are based on the following submitted article: [108] *S. Kernan Freire, T. He, C. Wang, E. Niforatos, and A. Bozzon. "Operators' Perspectives on Conversational AI for Knowledge Sharing: Challenges, Risks and Impact on Work". Submitted: Proceedings of the ACM on Human-Computer Interaction (CSCW 2024).*

CHAPTER 6: DISCUSSION AND CONCLUSION

The final chapter synthesizes the findings from the previous chapters by answering the main research questions, discussing implications for practitioners and researchers, and reflecting on the approach and future work. This chapter emphasizes the promising role of CAs in enhancing KS, thus improving the efficiency of problem-solving on the production floor. However, it also highlights the necessity for designs that prioritize human needs, organizational context, and ethical considerations, incorporating the perspectives of all stakeholders involved.

The chapter discusses the implications of our findings, such as the importance of maintaining high-quality knowledge and balancing the authoring burden on humans. Addressing privacy considerations and the tensions between operators and management are key to successful implementation. Reflections on our approach underscore the importance of engaging factory operators in the design process and the need for studies to investigate the long-term impacts.

Future work is outlined, focusing on refining LLM tools, conducting longitudinal impact studies, and exploring the transferability of our findings to other contexts. Developing frameworks for AI-powered KS to reconcile stakeholder values and incentivize engagement will also be important.

2

THE CONCEPT OF A COGNITIVE ASSISTANT FOR FACTORY OPERATORS

In this chapter, we present the design of a cognitive assistant to improve knowledge sharing among factory operators. Operating a complex and dynamic system, such as an agile manufacturing line, is a knowledge-intensive task. It imposes a steep learning curve on novice operators and prompts experienced operators to continuously discover new knowledge, share it, and retain it. In practice, training novices is resource-intensive, and the knowledge discovered by experts is not shared effectively. To tackle these challenges, we designed an AI system that provides cognitive augmentation to users of complex systems. We present a cognitive assistant that provides on-the-job training to novices while acquiring and sharing (tacit) knowledge from experts. Cognitive support is provided as dialectic recommendations for standard work instructions, decision-making, training material, and knowledge acquisition. These recommendations are adjusted to the user and context to minimize interruption and maximize relevance. The chapter describes the cognitive assistant design, how it interacts with users, its usage scenarios, and the challenges and opportunities in the factory context.

The content of this chapter is derived from the work published in S. Kernan Freire, S. S. Panicker, S. Ruiz-Arenas, Z. Rusák, and E. Niforatos. "A Cognitive Assistant for Operators: AI-Powered Knowledge Sharing on Complex Systems". In: *IEEE Pervasive Computing* (2022), pp. 1–9. DOI: 10.1109/MPRV.2022.3218600.

2.1. Introduction

Agile manufacturing is driven by the proliferation of product customization and on-demand production. Although it enables manufacturers to build high-standard, consumer-based products and respond to varying demands, it requires highly skilled personnel to operate agile production lines. To keep up with continuously changing production demands, cope with fluid product requirements and intense information processing, production-line operators need to adopt a working approach that welcomes change. Operators should be empowered to perform customization, allowing rapid adaptations, and providing operational flexibility while continuously enhancing their skills.

One of the challenges many industries face is the sharing of knowledge among operators, most notably tacit knowledge. Inherently, tacit knowledge is implicit and not codified, making it harder for individuals and firms to capture and assimilate [109]. Operators acquire tacit knowledge in their line of work to face the challenges of agile manufacturing. This tacit knowledge is hardly generalized, explained, or formalized for the standardization of manufacturing operations and for the training of novice operators. Although many definitions of tacit knowledge state that it cannot be verbally expressed, [110] consider that tacit knowledge can be subdivided into inexpressible tacit knowledge and expressible tacit knowledge. This categorization suggests that tacit knowledge has multiple dimensions in relation to the ability to articulate it into words, recipes and formulas, trade secrets, rules of thumb, and tricks [111].

Speed and effectiveness in training operators are another challenge in agile manufacturing, particularly for novice operators and when standard operating procedures are changed. Currently, factories rely heavily on word of mouth because paper and digital materials are difficult to organize and often become outdated. This results in high variations in operating quality, efficiency, and effectiveness. To tackle these challenges, we propose using a cognitive assistant, a conversational AI system for sharing knowledge. Thus, we pose the following dissertation research question: **What are the opportunities and design challenges when deploying cognitive assistants to support knowledge sharing between factory operators?** (RQ1). In this chapter, we introduce a mobile-based cognitive assistant that enables the transfer of expert knowledge to novice operators. Our cognitive assistant supports voice and text interaction and is accessed by a mobile device that operators can carry around the production line (Figure 2.1). Infrastructure stereoscopic cameras enable context-aware features (i.e., anonymous human tracking) to optimize the interaction with the cognitive assistant in a privacy-aware fashion.

We developed our solution in collaboration with multiple agile manufacturing partners, namely two detergent factories (The Netherlands and Italy), a fabric factory (Italy), and a fabric operator training school (Italy). The operators in detergent factories operate a production line comprised of numerous types of machines, from robot arms that stack boxes on a pallet to filling machines that inject detergent into canisters. The fabric operators operate loom machines that are comprised of two main parts: the weft and warp sides. The production activities in these factories are notorious for imposing a high cognitive workload on

Figure 2.1: An operator interacting with the cognitive assistant by voice and graphical user interface

their operators. An ever-changing selection must be produced that requires unique machine configurations, troubleshooting, and optimization techniques.

In the context of detergent production, even the characteristics of the components of a product can change abruptly (e.g., viscosity), requiring new machine settings and/or techniques. Operators work in shifts; they are remarkably busy, and existing knowledge-sharing methods (e.g., drafting a short report after an 8-hour shift) are time-consuming and error-prone. In such challenging manufacturing contexts, the ability to share new knowledge rapidly and accurately is essential yet non-trivial. The contributions of this chapter can be summarized as follows:

1. We present the system design of a cognitive assistant to acquire and share knowledge with factory operators.

2. We identify promising usage scenarios of cognitive assistants in supporting factory operators.

3. We provide a deeper understanding of the opportunities, challenges, and risks associated with using cognitive assistants by factory operators.

2.2. COGNITIVE ASSISTANTS FOR KNOWLEDGE SHARING

Since 1965, efforts have been made to acquire knowledge and store it in a Knowledge Base (KB) such that it could be used by AI systems to support human operators. Expert systems, or one of their modern counterparts, cognitive assistants, support humans in knowledge-intensive tasks, such as operating an agile manufacturing line. These AI systems are referred to as knowledge-based systems since they rely on a KB to support their users. Traditionally, the KBs on which these AI systems rely are static; they are defined by developers in collaboration with domain experts or scraped from existing databases. This is not surprising as the process of manually acquiring knowledge is resource intensive, a phenomenon known as the knowledge-acquisition bottleneck [54]. Using a static KB has various shortcomings: (1) knowledge changes over time; (2) the process of creating an exhaustive KB is resource-intensive; (3) it is difficult to update when gaps in the KB are identified; and (4) the KB misses important contextual information.

Recently, researchers have made considerable progress toward efficiently building and maintaining KBs for AI systems. For example, active learning from datasets [112], crowdsourcing the process [113], extracting knowledge from online forums [114], and interactively learning from (expert) users. All aforementioned solutions are promising, and all four could be integrated into a single system. However, knowledge is more than simple rules. As Davenport *et al.* [115] define it, "*knowledge is a fluid mix of framed experience, values, contextual information, and expert insights that provide a framework for evaluating and incorporating new experiences and information. It originates in and is applied in the minds of knowers.*" To acquire human knowledge using AI, some contextual information can be collected autonomously, such as the location of the operator (Figure 2.2); however, understanding the operators' reasoning process will be difficult without talking to them, for example, via a chatbot [116]. Interestingly, recent research has shown that factory operators provide much richer information, including explanations of their actions, when conversing with a voice assistant compared to sharing their knowledge on paper [117].

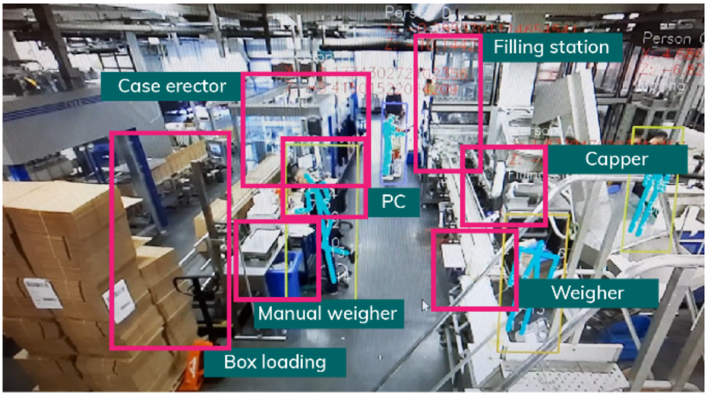

Figure 2.2: Stereoscopic camera view to track the operator's location and activity. Operators are tracked as 18-point skeleton frames thus protecting their privacy

Recent research on tacit knowledge in manufacturing has shown the immense value tacit knowledge can bring [118]. We now know that tacit and explicit knowledge exist on a continuum and that tacit knowledge can be converted into explicit knowledge [119]. Although our understanding of tacit knowledge has grown, it remains resource-intensive to effectively collect and share. In the manufacturing industry, researchers have manually extracted tacit knowledge using human motion capture, videos, and field interviews with experts and beginners [120]. However, this process requires skilled analysts to perform. Data analysis techniques can be used to identify tacit knowledge and store it in databases [121, 122]. Granted, these techniques do not provide operators with the opportunity to describe how or why they are doing things, nor do they offer a way to share their knowledge. To solve some of these challenges, a cognitive assistant could facilitate the sharing of tacit knowledge in a factory through dialectic interactions [123].

2.3. COGNITIVE ASSISTANT ARCHITECTURE

The cognitive assistant provides human-centered support to the operators on the production line; it "understands" how to collaborate effectively due to its context awareness, knowledge graph, user profile, and ability to continually learn (Figure 2.3). Context awareness is provided by a data stream from machines and stereoscopic cameras tracking the operator's location and activity. The Zed2 stereoscopic cameras[1] are positioned along the production line to provide full coverage. The stereolabs computer vision model[2] converts the camera stream into 18 XYZ points, providing an accurate and anonymous representation of the operator's skeleton (see Figure 2.2). To protect the operator's privacy, the live camera feed or skeleton data cannot be accessed by factory employees—only the cognitive assistant and researchers.

The live data streams from the machines provide the cognitive assistant with information on the machine's settings and status. This data is mapped to a knowledge layer that enables the cognitive assistant to identify relevant knowledge for users. Furthermore, by tracking context and production performance, it can predict when operators are using novel knowledge to reach higher performance (i.e., the best practices). In turn, the cognitive assistant can initiate an exploratory dialog with the operator to acquire (tacit) knowledge about these best practices. Any knowledge that is acquired is stored in the knowledge graph in relation to the context of its use. A knowledge graph represents a network of real-world entities (e.g., machine components, events, products) and their relationships (e.g., causes, is part of). This enables the cognitive assistant to recommend and explain the acquired knowledge if a comparable situation recurs. For example, if an operator would ask for help with a foaming product, the cognitive assistant might respond with the following information: "*I found the following causes for the described problem: high pressure, blocked nozzle and low product viscosity. I suggest the following solutions: reduce pressure and clean the nozzle filter. I found the following work instruction on the topic: <URL>*"

The cognitive assistant uses the following inputs for its recommendations: operator position, operator activity, knowledge graph, live machine data streams (e.g., machine settings and status). In turn, the cognitive assistant delivers the following output(s): task descriptions (e.g., work instruction steps), advice (e.g., suggesting potential causes or solutions to a problem), best practices (e.g., best settings for a product), media (e.g., video training material) and knowledge requests (e.g., why did the

[1] https://www.stereolabs.com/products/zed-2—last accessed November 20, 2024
[2] https://www.stereolabs.com/docs/body-tracking—last accessed November 20, 2024

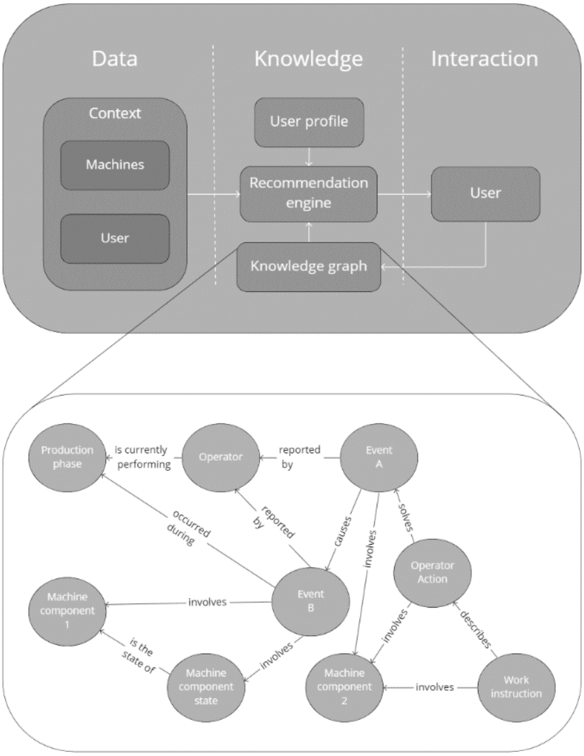

Figure 2.3: Cognitive assistant architecture, interactions, and a magnified view of its knowledge graph

operator change the settings).

2.3.1. Knowledge Representation

Knowledge representation (KR) methods organize relevant information in a specific domain in a structured way, allowing its use by intelligent systems, and reasoning programs [124]. Web-based services, software applications, robots, and human users make use of the represented knowledge through queries, to perform a task or provide a service. There are multiple types of knowledge representation methods. These methods can be categorized according to the type of knowledge they illustrate, namely *declarative*, *procedural*, *situational*, or *strategic* [125]. These categories can be mapped onto the explicit-to-inexpressible tacit knowledge spectrum. While explicit knowledge is already documented and easily expressible, inexpressible tacit knowledge acquisition requires the implementation of specialized knowledge-acquisition methods.

Declarative knowledge describes concepts and their semantic relationships. Machine component descriptions, issue description, root cause, and problem solution are examples of this type of knowledge in the manufacturing context. A lot of declarative knowledge can be extracted from existing factory documentation.

Procedural knowledge describes the procedures and methods to complete an action. Know-how on machine configuration and problem-solving procedures fits into this category of knowledge. Procedural knowledge contains expressible tacit knowledge, such as a description of steps, and inexpressible tacit knowledge, such as the "craft" operators develop through experience. Some explicit procedural knowledge may already be documented by factories and can be used by the cognitive assistant. The cognitive assistant can update existing procedural knowledge through dialogues with operators (e.g., by asking "*what did you do?*") and by observing their actions (e.g., operator tracking).

Situational knowledge is higher-level knowledge. It determines when and where to use the knowledge available in the specific domain of application. An example of this knowledge is the selection of methods, a solution, or a specific formula for problem solving. The ability of an operator to know when to use knowledge also contains expressible and inexpressible tacit knowledge. The cognitive assistant can acquire this knowledge through dialogues, data analysis, graph analysis and other AI techniques.

Strategic knowledge describes the reasons why a method or solution is selected. This last category may involve intuition and generalization, i.e., knowledge that is challenging for humans to explain and challenging for AI to acquire. In the manufacturing context, an operator uses strategic knowledge to adapt the solution for one problem to another similar problem. Strategic knowledge is strongly related to inexpressible tacit knowledge, therefore, the most challenging to acquire. The cognitive assistant can obtain this knowledge through exploratory dialogues (e.g., by asking why an operator did something), data analysis, graph analysis, and other AI techniques.

Our solution implements a knowledge graph (a declarative knowledge method) as an integrator element to connect procedural and situational knowledge methods in a coherent way. The knowledge graph describes and relates the main concepts associated to the manufacturing line through nodes and links, including production phase, machine components, product components, machine and product states, events, causes, and operator actions, among others (Figure 2.3). This representation allows connecting machine components with events, and actions to solve problems, enabling the cognitive assistant to navigate through the knowledge graph to get the most suitable recommendation for the operator. The "operator actions" nodes allow the user to process descriptions and the best practices describing the know-how of the operators. Action-type nodes provide procedural knowledge, while their relationships with events, and machine components derive situational knowledge needed to determine when to implement such a solution.

We use Neo4j[3], one of the leading graph database platforms, to host the KB that powers our cognitive assistant. Neo4j also allows for performing graph analytics, through the Neo4j Graph Data Science (GDS) library. GDS is a set of preloaded graph algorithms that contribute to the description of complex structures to unveil hidden

[3] https://neo4j.com/—last accessed November 20, 2024

patterns. Some of the analyses that can be performed through GDS are centrality, community detection, similarity, path-finding, prediction of topological links, and node embeddings. Centrality algorithms are used to determine the importance of different nodes in a graph. This algorithm can be implemented to identify the machine/product components that are prone to present more events (potential faults), the most popular solutions in the network, and the root causes that lead to more events, among others.

Pathfinding algorithms aim to find the shortest path between two or more nodes. This algorithm has great potential to determine the shortest procedure to solve an event and the shortest explanation of a cascading event. Community detection determines how groups of nodes are clustered and their tendency to break apart. This algorithm enables the cognitive assistant to identify interrelated events whose relationship has not yet been determined, as well as to gauge rare states and events (i.e., entities that are not connected to the rest of the network) that are depicted as "islands" on the knowledge graph.

Similarity algorithms are widely used in recommendation systems to assess the similarity between a pair of nodes. This analysis is used in the cognitive assistant context to extrapolate solutions and recommendations from one node to its pair, in cases in which the latter lacks information. The cognitive assistant can use the machine data streams to determine the overall performance production. A new best practice can be identified by comparing current performance with historical data. In turn, the cognitive assistant can collect any available data on best practice using context awareness and conduct a short interview with the operator to acquire relevant expressible tacit knowledge. In addition, it can use data analysis techniques to reveal inexpressible tacit knowledge.

2.4. Human-Assistant Interaction

The cognitive assistant's main user interface is a mobile app that supports interaction by voice, text (Figure 2.5), touch, graphically, and camera. This provides high adaptability; the voice interface can be used with a headset for hands-free and gaze-free interaction, and the user can take out the mobile device to look at an image or click on a link when necessary (see Figure 2.4).

The duration of the interactions is minimized, as the cognitive assistant can use its contextual awareness (operator location, activity, and machine data streams) to acquire important information and accelerate user interaction. This also facilitates intelligent conversation breakdown recovery. For example, if it is not clear which specific component the user is referring to, the cognitive assistant can infer the correct one based on the context (e.g., the operator's location).

2.4.1. Lean Information Exchange

The cognitive assistant suggests responses for the user based on the context and previous conversation steps. For example, when an operator starts to describe an issue, it suggests additional symptoms for the operator to check. In addition, it can use the operator's location to pre-fill information about the location(s) of an issue.

Figure 2.4: A conversation where an image of a loom part is displayed and then, the operator asks how to perform a fisherman's knot. The CA responds with a link to a video.

This feature aims to make dialogues more efficient by narrowing down possible conversation steps and encouraging the use of standardized vocabulary. Dialogues for training operators towards standardized terminology not only enables usage of synonyms of technical terms, but it also harmonizes vocabularies used by various stakeholders across the factory. The long-term goal is to train operators to provide high-quality data for further analysis, and to evaluate the progress of on-the-job training.

2.4.2. Lean Timing for Engaging Operators

Anonymous position and activity tracking enables the cognitive assistant to engage operators in issue reporting at a suitable time. Operator tracking (e.g., activity recognition) enables the cognitive assistant to detect opportune moments to report various types of issues and facilitate user interaction. The goal of this feature is to reduce the cognitive workload for operators by attempting to achieve balance between prompting the operator to recall necessary information and providing step-by-step documentation of issues in small snippets, parallel to performing issue handling actions. Additionally, it aims to reduce over-reliance on operators' organic memory to report an issue by supporting operators with dialogues that

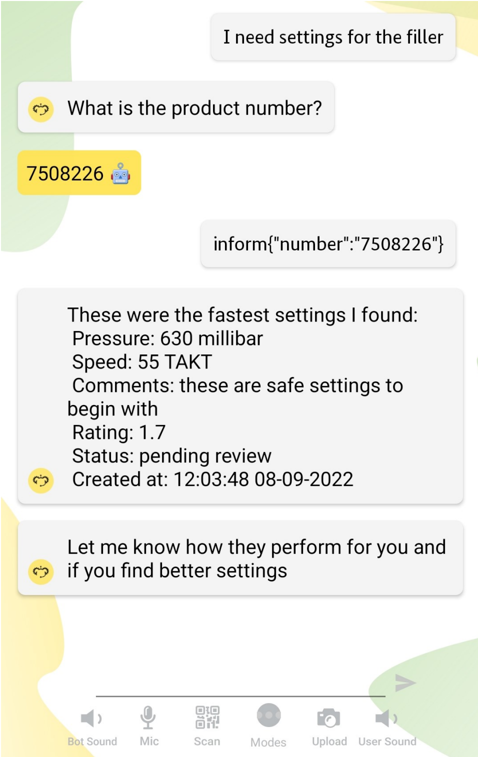

Figure 2.5: An detergent production line operator requests filling settings for a product

enable on-the-spot documentation, parallel to issue handling and production-line configuration activities.

2.4.3. Transparent Processing of User Suggestions

A context analysis we performed at the factories revealed that operators were unhappy about how their suggestions for improvements were handled as they did not receive any progress updates. The cognitive assistant provides the needed transparency by creating GitLab[4] support tickets for the operators. Consequently, operators can track the status of suggestions. This provides a central, transparent platform for operators to collectively track and discuss improvements to the knowledge base, the cognitive assistant, and the production line.

[4] https://about.gitlab.com—last accessed November 20, 2024

2.5. USAGE SCENARIOS

2.5.1. ACTIVE DOWNTIME

Operators are exceptionally busy keeping the production line running smoothly and handling issues. However, there are moments during a shift when the line is stopped and they must wait for other departments to perform tasks they depend on (e.g., technical support, quality control or logistics). During this period, the cognitive assistant can alert operators to recent updates and to standard operating procedures, recommend time-consuming training material, or provide an update on the knowledge they have provided (e.g., if it has been approved and how frequently it has been used). The cognitive assistant can keep track of what training the operator has already received to avoid repeated suggestions or identify when a refresher may be necessary.

2.5.2. MACHINE RECONFIGURATION

As operators reconfigure the production line, the cognitive assistant can provide recommendations relevant to the current step, for example, if the operator is standing next to the detergent filling machine during the clean phase, it can recommend the best practice for cleaning procedures or provide warnings regarding recent problems other operators have been facing. It can also answer simple questions the (novice) operators have about a task. If (novice) operators do not know how to request specific knowledge (e.g., they do not know the correct jargon), the cognitive assistant can aid them purely based on the context. The cognitive assistant will use the operator's position, activity, and machine status to provide a relevant recommendation.

2.5.3. ISSUE HANDLING

During issue handling, the cognitive assistant can provide suggestions for investigating actions, potential causes, and solutions to the issue. It can also collaborate with the operator to identify the root cause of a problem by using the 5-Whys approach, a proven agile manufacturing technique [126]. Even if the issue does not exactly match one in its knowledge base, it can use similarity models to find knowledge about similar situations that may be applicable. To prevent issues from occurring, the cognitive assistant can use machine learning models and acquired operator knowledge to alert operators in advance. Furthermore, the cognitive assistant can support the operators in efficient documentation of the issue and use the opportunity to acquire their knowledge.

2.5.4. SUGGESTIONS FOR IMPROVEMENTS

Operators can suggest improvements to the knowledge base, the interaction with the cognitive assistant, and the production line when it suits them. In addition, the cognitive assistant uses (machine learning) models to identify deviations from the best practices and applications of tacit knowledge. This enables the cognitive assistant to proactively acquire new best practices and, if necessary, interview the operators to capture their insights and thought processes.

2.6. Discussion

2.6.1. User Interaction

Interacting with the cognitive assistant by voice or text can be cumbersome and, to a certain extent, detrimental to actual operator performance. Input modality is paramount not only for operator's performance, but also for UX levels. In turn, satisfactory UX levels will be key to promoting the adoption of the cognitive assistant. Thus, the device that hosts the cognitive assistant becomes its embodiment and defines the interaction between the operator and the cognitive assistant. On the one hand, voice input is known to be error prone, and its performance will only deteriorate in noisy environments such as on a production line. On the other hand, typing text to interact with the cognitive assistant may not always be practical due to operator's equipment (e.g., gloves), or an activity that requires both hands to perform (e.g., adjusting the carousel torque while under the filler). In the detergent manufacturing environment, noise levels can reach 80 db(A) and operators are pressed for time. Therefore, we trained the cognitive assistant to understand keywords for several common features and suggest user responses as buttons to minimize interaction time and misunderstandings.

2.6.2. Knowledge Management

Factory management is interested in using cognitive assistants to identify best practices, share them and ensure that all operators follow them. However, expert operators are proud of their unique ways of operating and may resist standardization. Sometimes, during conversation, the experts omit important information, either, because they find it too obvious, are not aware of the importance of the omitted information, or intentionally, to keep competitive advantage, status, or power [127]. Typically, the management wishes to maintain control over what the best practice is, i.e., approve any new knowledge acquisitions. However, this might inhibit the speed of knowledge sharing. Peer assessment of new best practices using operator ratings could benchmark and accelerate the speed of knowledge sharing among operators. Peer assessment could also be used to maintain the accuracy of the knowledge base over time, as best practices will inevitably change. Another challenge is how to handle the cold start problem, namely, that the cognitive assistant needs to learn from operators before it can provide assistance. Understandably, users may reject the cognitive assistant if it was introduced as a blank slate. Therefore, we initiated the cognitive assistant with knowledge we manually acquired by interviewing operators and processing existing documentation.

2.6.3. Operator Safety

Activities performed on the production line, near or with heavy machinery, entail the inherent risk of serious injury or even death. Spontaneous interaction with the cognitive assistant could distract the operator and cause physical harm. We believe the context-aware component of the cognitive assistant can minimize the risk of operator injury. The cognitive assistant not only tracks the physical location of the operator on the production line, but also recognizes their activity (e.g., operating a hydraulic press). In the case of operating heavy equipment, the cognitive assistant should suppress its proactive prompting to prioritize the operator's safety over the potential to capture a new best practice. Here, detecting opportune moments to interact with the operator is not just a sought-after functionality, but an essential safety feature.

2.6.4. ETHICAL CONSIDERATIONS

An AI-powered system such as the cognitive assistant naturally raises many ethical concerns. For example, the long-term effects of an AI cognitive assistant on the operators, who use it daily for on-the-job training, have not been explored before. Here, we intend to help operators learn by forming a synergistic relationship between the operator and the cognitive assistant. However, we may be providing the "path of least resistance" to operators where they uncritically follow and over-rely on the instructions of the cognitive assistant. Learning-assessment tasks periodically delivered via the cognitive assistant (e.g., multiple choice questions) could help evaluate the actual expertise levels of the operators. The rich data streams necessary to the cognitive assistant can be misused by the management to intrude operators' privacy and violate their rights (e.g., excessive tracking). Data management plans, policies, and methods that protect and respect users' privacy and rights, respectively, are foundational in the dawning of Industry 5.0. Finally, it is important to consider how the authors of knowledge could be attributed and whether (monetary) incentive schemes are ethical.

2.6.5. LIMITATIONS

The cognitive assistant presented in this chapter was designed in collaboration with factory personnel, including operators, technicians and managers from three factories. However, as our access to the factories and the operators was arranged and mediated by the factory management, this may have affected how free the operators were to speak. Indeed, many of the assumptions that we based our design choices on, such as the performance difference in operators being caused by poor knowledge sharing practices, originate from management. Thus, the design and context we present in this chapter is skewed toward that of the management. In Chapter 5 we investigate the differences between the management and operator's perspectives in depth. Regarding the effectiveness of the system presented in this chapter, it remains to be tested with operators and managers. As such, this chapter represents a promising cognitive assistant design based on a context analysis of the agile manufacturing context.

2.7. Conclusion

We presented a cognitive assistant that supports agile manufacturing operators by orchestrating the sharing of (tacit) knowledge from experts to novices. Cognitive aid is provided through dialectic recommendations for standard work instructions, decision-making, training material, and knowledge acquisition. We emphasize the importance of designing cognitive assistants in alignment with human-centered principles, addressing the needs of operators for knowledge management, safety, and privacy. In Chapter 3, we will explore the effects of factors such as interaction modality and training on the user experience of cognitive assistant systems, developing more comprehensive design guidelines.

3

Lessons Learned from User Evaluations of a Cognitive Assistant

Building on the design for a cognitive assistant and usage scenarios presented in Chapter 2, we conducted an extensive user study with a functioning assistant system. We introduce CLAICA, a Continuously Learning AI Cognitive Assistant that supports factory operators of agile production lines. CLAICA learns from (experienced) operators, formalizes new knowledge, stores it in a knowledge base, along with contextual information, and shares it when relevant. We conducted a user study with 83 participants who performed eight knowledge exchange tasks with CLAICA, completed a survey, and provided qualitative feedback. Our results provide a deeper understanding of how prior training, context expertise, and interaction modality affect the user experience of cognitive assistants. Furthermore, we collected qualitative feedback on system usability, areas for improvement, challenges, risks, and potential impact on the shop floor. We draw on our results to elicit design and evaluation guidelines for cognitive assistants that support knowledge exchange in fast-paced and demanding environments, such as an agile production line.

The content of this chapter is derived from the work published in S. Kernan Freire, E. Niforatos, C. Wang, S. Ruiz-Arenas, M. Foosherian, S. Wellsandt, and A. Bozzon. “Lessons Learned from Designing and Evaluating CLAICA: A Continuously Learning AI Cognitive Assistant”. In: *Proceedings of the 28th International Conference on Intelligent User Interfaces.* IUI ’23. Sydney, NSW, Australia: Association for Computing Machinery, 2023, pp. 553–568. DOI: 10.1145/3581641.3584042.

3.1. Introduction

Operating a complex machine, such as a production line, can be a daunting task for an inexperienced operator, requiring extensive job training. Currently, training and knowledge sharing rely heavily on human interaction, where novices are often paired with experienced operators and receive a lot of one-on-one guidance for an extended period. However, this approach can be time-consuming and expensive, and expert operators may not have the capacity to share all relevant knowledge. Moreover, the departure of experienced operators or their retirement can result in a loss of valuable (tacit) knowledge, making it challenging to train new operators. Additionally, instruction materials is often inaccessible or outdated, inhibiting its use. As a result, training new operators is highly resource-intensive and requires a considerable amount of time [30].

Recently, intelligent products, systems and services are transforming how technicians share knowledge in manufacturing environments [19]. However, older systems are based primarily on predefined knowledge bases, which cost a significant amount of resources to develop and maintain [54]. Previous studies attempted to automatically uncover knowledge using NLP in existing maintenance reports, but many data quality problems were discovered [16]. Others came to the conclusion that technicians frequently provide informal descriptions of issues, which results in discrepancies and errors in the data; certain maintenance data, such as the actual root cause of a problem, are not always collected; and once the data is collected, it is often not used for a subsequent diagnosis [42]. Thus, the poor quality of reports inhibits (AI-facilitated) knowledge sharing between operators, requiring a more effective and user-friendly method of recording and sharing knowledge between factory operators [117].

We developed CLAICA (Continuously Learning Artificial Intelligence Cognitive Assistant) to support operators by providing on-the-job training and knowledge sharing. The CLAICA prototype can interact through a conversational user interface in a web browser across mobile and desktop devices. Its primary function is to learn best practices from (expert) operators and share this with others. In addition, it can recommend existing task instructions and training. The knowledge it acquires is stored in a continuously growing knowledge graph. It has several key advantages over a human mentor for novices; namely, (1) it is always available, (2) it will never forget information, and (3) it has direct and real-time access to production-line data. These advantages are critical in a modern agile production line, where a small number of human operate a complex connected system and many different products are produced, each of which requires specific setups. CLAICA's context awareness is powered by direct access to live production-line data (e.g., machine status, sensor data, error codes) that enables it to streamline interactions with users (e.g., it can detect what product is being produced without having to ask the user) and to help match the current situation to knowledge stored in its knowledge base.

To evaluate CLAICA's performance and its potential to enhance on-the-job training and knowledge sharing in factories, we conducted a comprehensive laboratory study involving 83 participants. We were guided by two primary objectives: firstly, to learn more about the effects of several factors on the user experience of the system, and secondly, to collect usability feedback from participants to learn more about the design requirements for cognitive assistants in factories. Expanding on the primary aim, our research question is: **How do modality, user training, and context experience affect the UX, usability, and interaction efficiency of cognitive assistants for knowledge sharing in factories?** (RQ2) The study participants performed eight knowledge exchange tasks with CLAICA related to representative manufacturing activities, such as requesting cognitive assistance to solve a specific problem and sharing new machine settings. The contributions of this Chapter can be summarized as follows:

1. We showcase how a cognitive assistant can exchange knowledge with operators about operating an agile production line.
2. We provide a deeper understanding on how context expertise, modality, and training affect the task performance, perceived workload, usability, and user experience of cognitive assistants.
3. We elicit requirements and design guidelines for future cognitive assistants beyond the manufacturing domain.

3.2. BACKGROUND

3.2.1. HUMAN-CENTERED MANUFACTURING

A key paradigm for modern production is the broad adoption of digital tools to increase, for example, productivity and sustainability. Thoben et al. concretized this paradigm with research challenges and application examples for Industry 4.0 and smart manufacturing [128]. The latter are two initiatives to systematically detail and implement this paradigm in manufacturing. While Industry 4.0 originated in Germany, smart manufacturing emerged in the United States. Countries such as Japan, Korea, and China created similar initiatives. These initiatives are largely technology-focused and do not adequately account for human needs, involvement, and collaboration with AI systems [28]. The next paradigm addresses this shortcoming through human-centric manufacturing, called Industry 5.0 [28, 29]. Müller reports the findings of an expert group regarding technologies contributing to more human-centric manufacturing [129]. They include, among others, individualized human-machine interaction technologies accounting for the strengths of humans and machines. Also, using AI to assist humans in understanding causalities in complex and dynamic systems. Romero et al. introduced similar ideas stressing the importance of socially sustainable manufacturing [130]. They proposed the vision of Operator 4.0, which focuses on trusting and interaction-based relationships between humans and machines. Eight types of operators illustrate how technology could enhance human capabilities and skills.

Meanwhile, agile manufacturing enables product customization and on-demand production to respond quickly to customer needs and market changes. Moreover, because of the low cost of Internet of Things (IoT), IoT devices have gained popularity in manufacturing settings to optimize business workflows and processes, improve safety and improve research and development. Nevertheless, such devices also generate unprecedented volumes of sensor data, necessitating operators' constant attention and dramatically increasing the incurred cognitive load. Interestingly, working in an agile connected production line is a real challenge for inexperienced operators. Even for experienced operators, operating an agile production line is a knowledge-intensive task that requires enormous cognitive resources [30]. For example, a single operator may need to operate and fix numerous machines along a production line simultaneously while also (re)optimizing the setup for more than 100 different products. Therefore, systems that could reduce cognitive demands and allow quick feedback are needed to effectively support operators personally.

3.2.2. Cognitive Assistants

Unlike systems designed to replace humans in specific tasks (e.g., industrial robots), cognitive assistants strive to complement human abilities to accomplish complex tasks, such as aiding life-long education and machine operation [131–133]. In addition, such assistants often outperform human capacities for communication and memory in various ways, such as simultaneously providing dependable and repeatable communication between numerous users [131, 132]. To achieve the aims mentioned above, cognitive assistants should support efficient human-machine interfacing via natural language processing, interpretation of gestures, perception, vision, and sounds, augmented reality to provide additional layers of information, and others [131, 132]

Of these, the most widely used interaction method for cognitive assistants is conversational agent-based natural language communication, which involves natural language understanding, generation, and dialog for implementation [131]. Conversational agents engage with people using natural language, which could perform labor-intensive jobs at low cost in a variety of industries, such as customer service, healthcare, education, e-banking, and personal assistants [84]. For example, recent work has proven conversational agents to be an effective educational tool for communicating breast cancer risk and suggested medical guidelines to women, leading to a significant increase in breast cancer genetics knowledge [134]. Furthermore, advances in context awareness make it possible to improve a conversational agent's usefulness. For example, acquiring more accurate knowledge about city locations by asking questions when users are there [77] or inferring context from user utterances to provide more relevant tourist recommendations [135]. A virtual assistant called Amber was proposed by Kimani et al. using a sensing framework that could record users' faces, speech, and app usage to aid users with job prioritization, provide reminders, and inhibit social media diversions [136].

Conversational agent-based cognitive assistants have also shown promise in supporting one's cognitive processing. [133] developed LIZA, a cognitive assistant to enhance users' reasoning and decision-making abilities. By holding a conversation with test subjects to help them solve common heuristic and bias problems, LIZA helped test subjects to improve their reasoning skills and achieve significantly higher learning gain [133]. Numerous application scenarios have also been proposed in which cognitive assistants have positive cognitive effects. From education and training, such as by employing lifelong learning to retrain adult operators to meet shifting technological demands, to elderly care by facilitating interaction with those suffering by cognitive decline [131].

3.2.3. COGNITIVE ASSISTANTS IN MANUFACTURING

Industrial applications for AI assistants—similar to Alexa[1], Google Assistant[2], or Siri[3]—are an emerging research topic. These prototypes emerged in different research communities with different names (e.g., intelligent (virtual/personal) assistants, digital assistants, software robots, or simply chatbots). AI assistants in manufacturing can bear significant benefits [137]. These include, for instance, central access to heterogeneous information systems, the delegation of tasks, and gaze-free and hands-free interactions during work. Furthermore, AI assistants can be used for cognitive support, such as training operators [102] and adjusting machine parameters [138]. When used to support cognitive processes such as learning and reasoning, we refer to cognitive assistants.

Most of the literature regarding cognitive assistants in manufacturing focuses on knowledge and information delivery, for example, context-aware assistance (e.g., [139]), recommendations for predictive maintenance (e.g., [140]), decision support based on business analytics of shop-floor data (e.g., [141, 142]). For instance, Rodriguez *et al.* [143] present a mixed reality assistance system to support real-time assembly operations. They evaluated the operation context through a recognition system that determines the completion of each assembly step to derive the next instruction. Belkadi *et al.* [144] proposes a context-aware knowledge-based system aiming to support manufacturing operators. They integrate knowledge management, context management, and simulation management modules to support the decision making of the operators in real-time. As for Rodriguez *et al.* [143], context management gets contextual information to understand the current user's situation and implements simulation techniques to anticipate the effect of the operator decisions. Büttner *et al.* [145] implemented a hand-tracking algorithm to identify wrong-picking actions and errors in the assembly process. Tao *et al.* [61] implements wearable device sensing and environmental sensing to capture operator activity in the workplace to guide them in the execution of their tasks, and Josifovska *et al.* [146] integrates a context manager module which includes a Digital twin of a human that simulates specific human abilities and preferences to enable assistance system adaptation. Longo *et al.* [147] demonstrated an AI assistant integrated into an augmented reality application to train machine operators. Their prototype provides information about safety measures, potential hazards, machine status and operations, and quality control procedures. Besides, it instructs users on lubrication, greasing, cleaning, checking, and restoring hydraulic pressure or fluids and lube for maintenance.

Researchers have used several approaches to evaluate the effect of instruction delivery on operators' mental workload and its effectiveness. Funk *et al.*

[1] https://www.alexa.com/about—last accessed November 20, 2024
[2] https://assistant.google.com/—last accessed November 20, 2024
[3] https://www.apple.com/siri/—last accessed November 20, 2024

[139] evaluated the workload effect associated with the delivery of instructions on assembly work by monitoring biosignals (such as heart rate, galvanic skin response, electroencephalography, and electromyography) and indicators such as task completion time and error rates. Likewise, Kosch *et al.* [148] evaluated the cognitive workload produced by in situ projections during the execution of manual assembly tasks. They implemented electroencephalography (EEG) to monitor cognitive workload and compared the results with those obtained by traditional paper-based instructions. In line with the above-mentioned approaches, Funk *et al.* [149] proposed a standardized experiment design for evaluating the effect of interactive instructions in heterogeneous assembly tasks.

Existing literature has also explored acquiring knowledge from operators, but to a significantly lesser extent. This could be partially attributed to the less immediate benefits; however, we argue that the ability to continuously learn is necessary for the long-term success of AI assistants on the shop floor. Fenoglio *et al.* [75] propose a system for capturing explicit and tacit knowledge (through best practices) from experienced operators in industrial domains. They implement a role-playing game where a virtual agent interacts with human experts and knowledge engineers to extract and represent knowledge in an iterative way. Despite this system providing means to capture tacit knowledge, it requires the intervention of a human agent, as they argue that it is impossible to capture tacit knowledge with an algorithmic technique. Likewise, Soliman and Vanharanta [150] suggests a model for knowledge creation and retention through artificial intelligence. However, there is no practical application of the model reported in literature, and Hoerner *et al.* [58] propose a digital assistance system to support operator troubleshooting processes on the shop floor. For this purpose, a method to capture and structure expert's tacit knowledge was developed. However, this method is not executed by the digital assistant. It is performed by human experts who are in charge of extracting and representing the obtained knowledge and delivering it as input to the system.

However, and to the best of our knowledge, no existing conversational AI assistant captures (tacit) knowledge from experienced operators with the purpose of structuring it, storing it, and re-sharing it with novices in real-time and on the shop floor. CLAICA eases the learning curve for novice operators by serving as a dialectic mediator among novices and experienced operators. Even for experienced personnel, CLAICA provides multiple opportunities to rehearse and test acquired knowledge and skills, while constantly discovering and formalizing new knowledge or knowledge that has been overlooked (e.g., tacit knowledge).

3.3. CLAICA: Continuously Learning AI Cognitive Assistant

The primary goal of CLAICA is to enable knowledge exchange between shop floor operators. It continuously learns by acquiring knowledge from operators through dialectic interactions, allowing it to efficiently share up-to-date knowledge. In addition, it can recommend existing work instructions and perform information retrieval tasks. Furthermore, operators can provide feedback on the knowledge they receive to improve recommendations over time. What sets CLAICA apart from the state-of-the-art is its ability to efficiently acquire knowledge from operators on the shop floor without human involvement and store it in a knowledge graph along with contextual information. However, knowledge managers may still be needed to perform quality control by approving, reviewing, and removing elements of knowledge. Ultimately, CLAICA aims to reduce the burden on "knowledge engineers" and improve knowledge sharing.

3.3.1. CO-DESIGNING CLAICA

CLAICA was developed in close collaboration with an industrial company, a detergent producer. Their ambitions for the assistant are to result in faster training of new operators and higher OEE (Overall Equipment Effectiveness) [151], that is, the percentage of manufacturing time that is truly productive. During the early design phase of CLAICA, we conducted semi-structured interviews and focus groups with operators and management at two detergent factories. We explored three main topics, namely what opportunities are there to support machine operators with a cognitive assistant, what wishes do factory employees have regarding the (interactive) capabilities of the assistant, and what challenges the assistant might face from a user acceptance perspective. We interviewed six machine operators, two maintenance technicians, two shift leaders, four engineers, and one factory director. We pinpointed the following opportunities: (1) identify the best practices of operators, elicit this knowledge, and share it with others, (2) make existing work instructions more easily accessible, (3) help operators identify the root cause of problems, (4) provide operators with access to machine data anywhere on the production line, and (5) create high-quality issue reports. The operators were primarily interested in receiving ubiquitous access to machine data, while their supervisors were also concerned with facilitating knowledge sharing among operators, providing access to instructions, and creating better issue reports. The shift leaders suggested that operators, even experts, could benefit from suggestions; however, the operators themselves disagreed. We noted that the operators were very proud of their skills and proud that they did not need instruction material. However, they thought that novice operators could benefit from easy access to up-to-date documentation. Existing documentation resides in paper and digital form, but is poorly structured and frequently outdated. As such, novice operators are trained almost exclusively by experienced operators on-the-job. Due to the complexities of operating an agile production line, this process is lengthy and, therefore, costly. Furthermore, it is risky from the company's perspective, as a lot of valuable (tacit) knowledge will be lost when experienced operators leave.

We collected 100 issue descriptions on one of the production lines over three days. We asked the operators to verbally describe the location, symptoms, and cause of each issue as it occurred. The operators used a mono headset during data collection (see Figure 3.1).

We analyzed the resulting reports to identify opportunities and challenges for CLAICA. We observed that operators used different terms and acronyms when describing the same machines (e.g., depa, palletizer, depalletizer). In addition, the

Figure 3.1: An operator describing a problem with the production line

production line breaks down frequently (up to 30 times in an eight-hour shift), and the operators are under intensive time pressure. As a result, many problems go undocumented. Once we built a working prototype, we presented it to the operators (see Figure 3.2) to elicit further feedback.

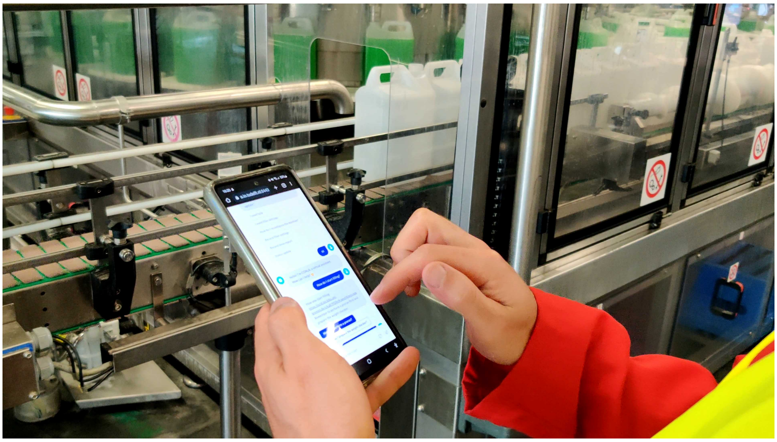

Figure 3.2: An operator using CLAICA at the production line

The insights from the interviews, the focus groups, and the collection of issue descriptions resulted in the following design requirements for CLAICA.

- It must provide accurate cognitive support to novices
- It must interact efficiently and reliably

- It must support user feedback
- It must continuously learn from its (expert) users
- There must be incentives for (expert) users to share their knowledge (this responsibility is shared with the company)
- It must reduce the reliance on expert operators for training and supporting novices
- It must be transparent about what data it collects and how it provides recommendations
- It must be able to handle divergent phrasing and misunderstandings gracefully

3

3.3.2. SYSTEM DESIGN

CAPABILITIES

CLAICA employs a conversational user interface (CUI) as its primary means of interacting with users, a knowledge graph for storing its knowledge, and a cognitive engine for processing and integrating all the information streams (see Figure 3.3). CLAICA can collect information about the context autonomously (e.g., machine states). As such, the context awareness provided by the live production data enables CLAICA to streamline interaction, minimize its duration, and input burden (e.g., by auto-filling some of the information for the user).

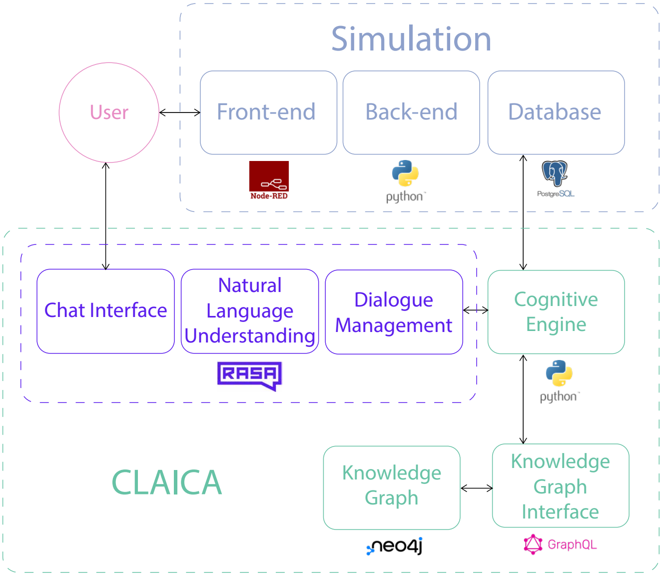

Figure 3.3: CLAICA's architecture (below) and the simulation (above)

CLAICA can continuously learn from operators in the following scenarios.

1. *A operator shares how they diagnosed and solved an issue.* Upon classifying the operator's intent to provide an issue report, CLAICA will start requesting information from the user. Each issue report is split into three sections, namely symptoms, cause, and solution. CLAICA asks the operator to provide a description of each of these in text. While processing the operator's response, CLAICA attempts to extract several named entities such as any machine components (e.g., nozzle 23), machine component states (e.g., overheated), product components (e.g., label) and product component states (e.g., crooked), error codes, and operator actions. If CLAICA is unable to extract at least one machine component associated with the symptom of the issue, it asks the operator to specify one manually. Furthermore, it uses its context awareness to suggest additional entries as buttons in the CUI, such as machine error codes. Finally, the operators' descriptions, extracted entities, and context information are stored in the knowledge graph.

2. *A operator shares machine settings for a product.* In this scenario, CLAICA uses its machine integration to collect some of the information automatically, for example, which product is being produced and with what settings. However, it still asks the operator for confirmation and the option to specify something manually. Additionally, CLAICA asks the user for comments on the settings, for example, if they are safe or risky. In turn, this information is stored in the knowledge graph.

3. *A operator rates existing machine settings.* Upon receiving recommended machine settings from CLAICA, operators are encouraged to rate them from 1 to 5. CLAICA takes these ratings into account when recommending the settings to other works. In addition, a rating of 1 will automatically triger the settings to be retracted and flagged for review by a knowledge manager.

In addition, CLAICA can provide knowledge in the following scenarios.

1. *A operator asks for help.* When a operator asks for help, CLAICA asks them to describe the situation. It attempts to extract several named entities, such as machine component(s) and their states. In turn, it uses these entities to search its knowledge graph for the most relevant existing solutions and/or documents and recommends these to the operator.

2. *A operator asks for machine settings for a product.* When the user requests machine settings, CLAICA checks which product is being produced and asks for user confirmation. Then, it searches its knowledge base for the highest-rated settings for that product and presents it (see Figure 3.4).

3. *A operator asks how to perform a specific task.* CLAICA's knowledge base also includes existing documentation on how to perform several tasks. The operator can ask for them explicitly, for example, "how do I perform a prerun?" (see Figure 3.5).

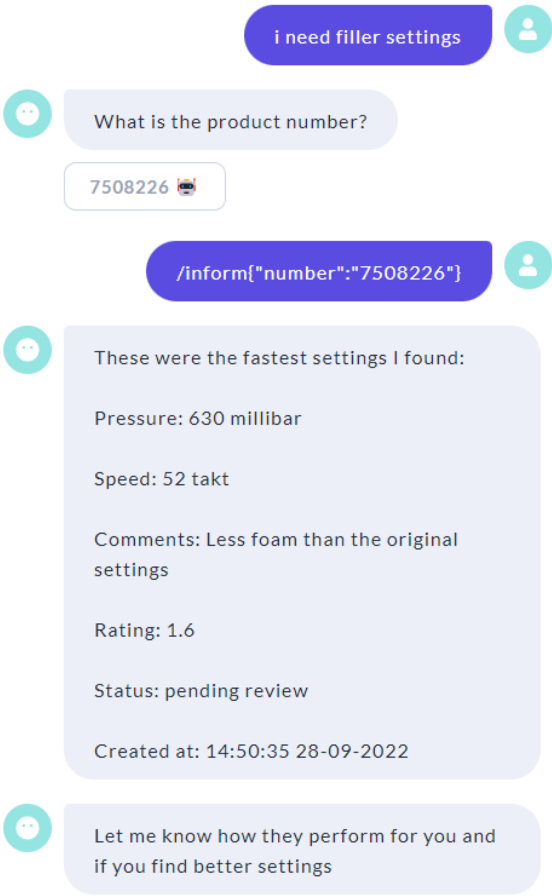

Figure 3.4: A user requests filling machine settings for a product

ASSISTANT FRAMEWORK

We use an open-source conversational AI assistant framework, Rasa[4], to build CLAICA (see Figure 3.3. We use Rasa due to its flexibility and performance; it can easily be customized for specific use cases and connected to additional components, such as knowledge bases. This includes adding custom Python scripts (Cognitive Engine), adjusting the NLP pipeline, and using custom entity extractors, such as Duckling[5] (e.g., to extract pressure values from user utterances). The NLU, dialogue management uses several base features that have universally applicable intents and conversation patterns, for example, greeting and restarting the assistant. Additionally, we have added features that contain specific training data, domain descriptions, and custom actions for CLAICA. The Cognitive Engine, also known as the custom actions

[4] https://rasa.com/—last accessed November 20, 2024
[5] https://duckling.wit.ai/—last accessed November 20, 2024

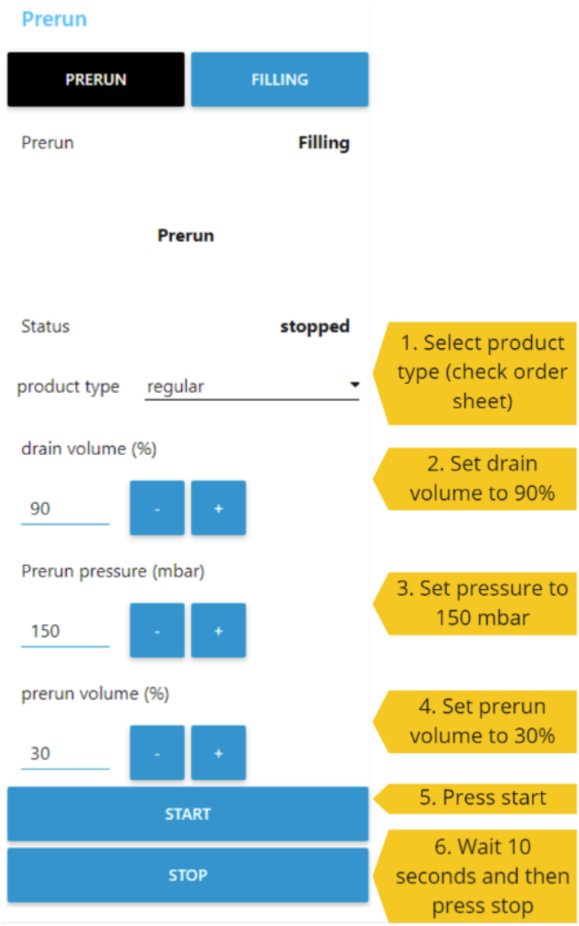

Figure 3.5: Task instructions for a prerun

server, is the place where all incoming streams of information, such as machine data, user-provided information, and knowledge base, are handled, analyzed, and responded to. It also contains the logic to ensure all the relevant information is collected from user's during knowledge acquisition (e.g., during an issue, CLAICA needs to collect the machine component ID, symptom, cause and applied solution). We run a Rasa X server that supports a browser-based chat interface (see Figure 3.4). It also provides a browser-based interface to review user conversations and apply improvements.

Continuously Growing Knowledge Base

The knowledge base is a Neo4j[6] knowledge graph, also known as a semantic web. It can be queried by the assistant using GraphQL[7] resolvers. A knowledge graph is a type of data representation that uses nodes and edges to map relationships and information between entities. It helps machines, like CLAICA, to "understand" the meaning and context of data, providing a more structured and interconnected view of information. For CLAICA, we define a description of an issue symptom as an "event" node that will have a "caused by" relationship to another event that describes the cause. In turn, the symptoms event will have a "solved by" relationship to a solution node. Each of the event nodes can have relationships with other entities, such as machine components and their states. This gives CLAICA a robust way to find relevant information (e.g., a solution), to a situation described by the operator.

Simulating the Production Line

Testing an intervention, such as a cognitive assistant, in a detergent factory and in situ is challenging due to the potential for dangerous and costly operating errors. Problems can arise from incorrect recommendations and the cognitive load or distractions imposed by interactions with the assistant. Furthermore, since the target audience for the assistant is new employees, the pool of available test participants is small and unreliable. Therefore, we created a detergent factory simulation to evaluate the assistant in the lab before introducing it in the wild. The simulation features several GUIs that connect to a model in the back-end. These include GUIs where users can control the most important machines on the production line, namely the detergent container filling machine (Figs. 3.6) and the detergent container weight checker machine (Fig. 3.7).

The front-end was built using Node-RED[8], which connects to a PostgreSQL[9] database and machine models written in Python[10]. It supports several simple tasks related to the preparation of machines for production, as well as more complex optimization and problem-solving tasks. The graphical user interfaces (GUIs) are browser-based and designed to be accessed on a tablet or laptop. To simulate a real factory scenario as closely as possible, we examined the set-up found in a detergent factory and modeled our lab set-up on it. For example, for the GUI, we matched the information displayed, the user controls, their physical height, and the distance between them. The production line simulator was validated through several focus group sessions with factory employees ranging from shop floor operators to process improvement engineers.

[6] https://neo4j.com/—last accessed November 20, 2024
[7] https://graphql.org/—last accessed November 20, 2024
[8] https://nodered.org/—last accessed November 20, 2024
[9] https://www.postgresql.org/—last accessed November 20, 2024
[10] https://www.python.org/—last accessed November 20, 2024

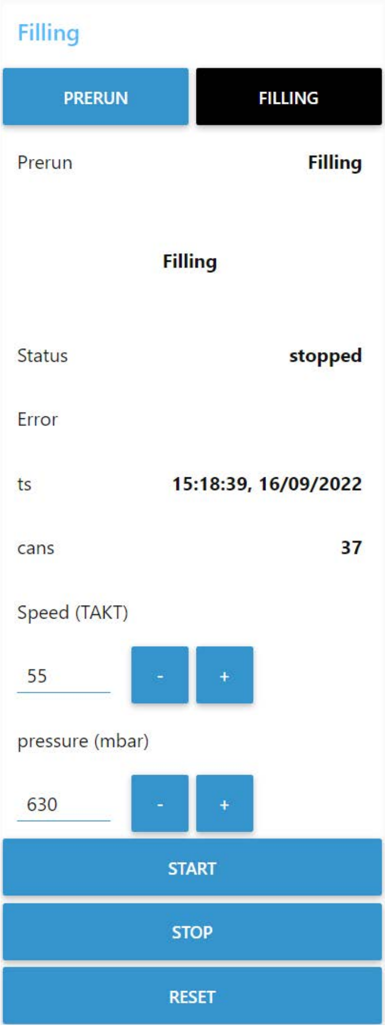

Figure 3.6: The simulated user interface for the filling machine

3.4. STUDY AND METHODOLOGY

We conducted a between-subjects user study with 83 participants to evaluate user performance, usability, user experience and perceived workload when interacting with CLAICA. The user study took approximately one hour and consisted of the following three parts: a demonstration of the (simulated) production line, a demonstration of the assistant, and a user test. The user test involved performing eight knowledge exchange tasks with CLAICA. These included information retrieval (e.g., “Find instructions on how to perform a prerun”), knowledge sharing, and requesting cognitive support. Note that the participants only interacted with the

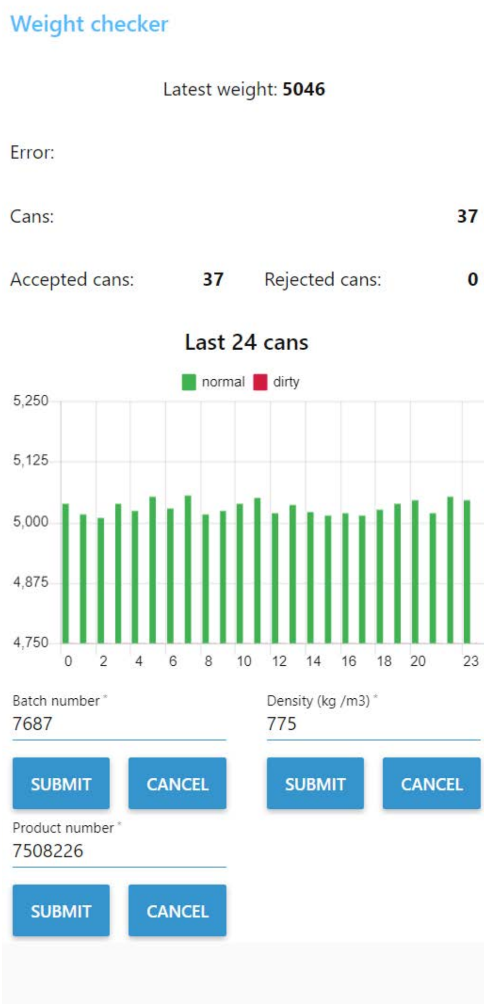

Figure 3.7: The simulated user interface for the weight checker machine

assistant; they did not have to perform any actions with the (simulated) production line. Participants then completed a survey, including a NASA Task Load Index questionnaire (NASA-TLX) [152], a User Experience Questionnaire (UEQ) [153], and a System Usabilty Scale (SUS) questionnaire [154]. At the end of the study, we posed several open-ended questions to initiate a discussion and obtain qualitative feedback from our participants.

3.4.1. RESEARCH QUESTIONS

In this chapter, we sought to answer the following dissertation research question: **How do modality, user training, and context experience affect the UX, usability,**

and interaction efficiency of CAs for knowledge sharing in factories?(see RQ2). This research question for this chapter is broken down into three parts and discussed below.

RQ2.1 *How does prior training affect user performance, perceived workload, user experience and perceived usability when using CLAICA?*
When introducing a new tool or system like CLAICA to end users, its introduction will probably include some training on what it can do and how to use it. Managers would like to set expectations and manage the change process with care. To ensure adoption, managers should highlight the perceived usefulness of a system, especially the improvements that it may bring to efficiency [155]. Furthermore, training could help users become more effective with the new system. We plan to provide training to end users when CLAICA is deployed in situ; therefore, we would like to evaluate the effect of providing training or not ("trained" vs. "untrained" participants). As we expect users will need time to learn how to interact with CLAICA, we think that the trained group will be able to interact more efficiently with CLAICA and, therefore, have a better use experience.

RQ2.2 *How does context expertise affect user performance, perceived workload, user experience and perceived usability when using CLAICA?*
Due to the challenges of conducting studies with real end-users, many researchers opt to recruit participants locally (e.g., students or colleagues). The same is true for companies developing new products - access to end users can be difficult for many reasons, for example, confidentiality [156]. Therefore, it is valuable to understand how context expertise affects the user experience of AI assistants such as CLAICA ("operator" vs. "layman" participants). We do not expect context expertise will help users interact with CLAICA more efficiently; however, we think that it may affect the subjective user experience as they understand the implications of CLAICA in the context better.

RQ2.3 *How does the modality of interaction with CLAICA affect user performance, perceived workload, user experience and perceived usability?*
From the formative study interviews with end users, we learn that they currently use a fixed computer to access instructions and create issue reports. However, the production line can be long, so having to walk back to the computer takes time and may be a barrier to its use. Furthermore, they may not remember all the details of an instruction when they return to the relevant machine on the production line. Regardless, the laptop's larger screen and keyboard might help users retrieve and provide information faster and we use it as a baseline to compare against ("smartphone" vs. "laptop"). Previous research has shown that typing on computers is faster, but that smartphones can benefit more from suggested words [157]. Therefore, we expect that the laptop group will be able to perform tasks faster and have a better user experience.

3.4.2. Participants

We recruited 83 participants (44 male, 34 female, three "other" and four "preferred not to say") with age ranges from 18 to 64. Most of the participants ($N = 50$) were 17–29 years old, followed by the 30–39 age bracket ($N = 16$). All participants were able to communicate clearly in written and spoken English. The participants were recruited from the following four groups: factory employees ($N = 12$), external Human-Computer Interaction (HCI) researchers ($N = 16$), local HCI researchers ($N = 18$) and master's students ($N = 37$).

3.4.3. Procedure

Before commencing a user trial, participants were asked to read and sign an informed consent form that had been approved by an ethical board. Participants were randomly but equally split between the "smartphone" group and the "laptop" group. Next, participants were shown a short video introduction to the context of manufacturing and AI assistants. We then presented a three-minute video to introduce the (simulated) production line. The first part of the video introduced the detergent production line, a few of the operator's primary tasks and the machines we included in the simulation (filling machine and weight checker). Then, participants were shown how we simulated the filling machine (see Figure 3.6) and the weight checker (see Figure 3.7). Following the first video, the participants were given three minutes to write down comments on the realism of the simulation and how it could be used and improved. Participants were asked to write their comments as digital post-it notes. Then, the participants from the "trained" group were shown a five-minute training video demonstrating the capabilities of the assistant and how to use it. Again, they were given three minutes to write their comments. The "untrained" group, who performed the study at another time and place, was asked to write down comments on a related topic. Therefore, we ensure a comparable total experiment time and workload. The participants in the "smartphone" group were asked to interact with CLAICA using a web browser on their smartphone, while the participants in the "laptop" group used a web browser on their laptop. Then 10 minutes were given to complete eight assigned knowledge exchange tasks with CLAICA. Participants were asked to interact naturally in English. Immediately after completion, they were instructed to complete several surveys, namely NASA-TLX, SUS, UEQ, demographics (age, gender, occupation, years of experience in occupation), and personal factors related to chatbot experience. After completing the survey, we moderated a 10-minute discussion on their experience using the assistant and opportunities for improvements.

3.4.4. Measures

We measured the following quantitative objective variables: Task completion time, Task completion rate (yes/no), Conversation turns between user and CLAICA, User utterance length (words), Conversation breakdowns (when CLAICA cannot classify the user's intent), and User typing errors. The task completion time was automatically calculated by the Qualtrics[11] survey using the time difference between the first and last click on the task instruction page. The chat logs were extracted from an SQL Tracker Store [12] using SQL queries and a Python script. We then semi-automatically extracted participant IDs to match the chats to the survey responses. Quantitative objective variables related to the interactions were automatically calculated using a Python script. We used the Enchant [13] library to identify user typos and checked the assistant's failure response "/restart" to count conversation breakdowns. For the sake of brevity, we omit these metrics in this chapter.

We measured the following quantitative subjective variables: workload using the NASA-TLX, usability using the SUS, and user experience using the UEQ. We collected the following demographics and personal factors: Age, Gender, Occupation, Years of experience in occupation. In addition, we collected the following self-reported factors related to chatbots and technology: prior experience with chatbots and tech-savviness. We asked the following questions with likert scales as a response: "*I am familiar with chatbot technologies*," "*I use chatbots frequently*," "*I consider myself an advanced technology user*," and "*I am eager to try new technologies*," and "*I am good at solving technical problems*." The first four of these questions have been used in prior work on chatbot breakdown strategies [158, 159].

3.4.5. Qualitative Analysis

The last part of the user study consisted of asking the participants several open-ended questions to gather detailed insights from users about their experience executing tasks with the assistant. The questions were as follows: "How did you experience the assistant?", "What worked and what didn't?" and "How could it be improved?". After completing the tasks and filling our the survey, the participants were given 5 minutes to write digital post-it notes on a Miro board [14]. A total of 148 text entries were collected. We performed a content analysis to investigate the user responses. [160] define content analysis as "*the process of developing a representative description of text or other unstructured input*". We used two subjective coders, who read the data multiple times and then followed a mixture of emergent coding and a priori coding to create the key categories. The first and second level are based on a priori coding to clean and sub-categorize the data according to this study's goals. Comments were not processed further if they were not directly relevant to answer the research questions.

3.5. Results

To decide on our statistical methods, we first performed all the necessary pre-tests, such as Shapiro-Wilk tests of normality and Levene's tests of homogeneity of variance. We omit the pre-tests for brevity. Depending on the statistical test at hand, we report averages and standard deviations (parametric), or median values (non-parametric).

[11] https://www.qualtrics.com/—last accessed November 20, 2024
[12] https://rasa.com/docs/rasa/tracker-stores/—last accessed November 20, 2024
[13] https://abiword.github.io/enchant/—last accessed November 20, 2024
[14] https://miro.com/whiteboard/—last accessed November 20, 2024

We started our analyses by computing correlations between personal factors (e.g., demographics and prior experience with chatbots), as they can confound the relationship between the independent and dependent variables. We found the following two significant correlations: age is positively correlated with SUS score ($r(81) = .238, p < .05$) and self-reported technical problem-solving skills are positively correlated with completion rate ($r(83) = .308, p < .05$). These correlations indicate that (a) the older our participants were, the higher their usability ratings about CLAICA, and (b) the higher the self-reported problem-solving skills, the more probable it is for our participants to successfully complete a knowledge exchange task with CLAICA. We then investigated if there were any significant differences between the median age and technical problem-solving skills of the groups. A Mann-Whitney U test showed that there is a significant difference in the median age category between the layman group ($Mdn = 1$) and operator group ($Mdn = 4 \ /mid \ U = 80.500, p < .001$), between the untrained group ($Mdn = 2$) and the trained group ($Mdn = 1 \mid U = 311.500, p < .05$), but no significant difference between the median age category of the smartphone group ($Mdn = 1$) and laptop group ($Mdn = 1 \mid U = 775.500, p = .746$). Regarding self-reported technical problem-solving skills, a series of Mann-Whitney U tests did not reveal any significant differences between the laymen ($Mdn = 4$) and the operators ($Mdn = 4$)($U = 379, p = .519$), or between the untrained ($Mdn = 4$) and the trained group ($Mdn = 4$)($U = 391.500, p = .336$), or between the smartphone ($Mdn = 4$) and the laptop group ($Mdn = 4$)($U = 791.500, p = .596$).

3.5.1. EFFECTS OF PRIOR TRAINING

At first, we investigated whether receiving instructions (training) on what a cognitive assistant can do would have any effect on user performance, expressed as overall task completion times and task completion rates. However, a non-parametric Mann-Whitney U test displayed no significant difference in the median overall task completion times between untrained ($Mdn = 423.463$) and trained participants ($Mdn = 370.165 \mid U = 302, p = .099$). Similarly, a non-parametric Mann-Whitney U test displayed no significant difference in the median overall task completion rate between untrained ($Mdn = 6.5$) and trained participants ($Mdn = 7 \mid U = 389.500, p = .332$). These findings indicate that **prior training on how to use CLAICA did not significantly affect task performance when using CLAICA (RQ2.1).**

Next, we investigated if prior training on CLAICA impacts perceived workload (NASA-TLX) and usability (SUS). An independent-samples t-test revealed a significant difference in the average NASA-TLX scores between the untrained ($M = 35.651, SD = 18.261$) and the trained group ($M = 46.058, SD = 13.080 \mid t(69) = -2.590, p < .05$)(see Figure 3.8).

Similarly, an independent-samples t-test revealed also a significant difference in the average SUS scores between the untrained ($M = 59.559, SD = 19.946$) and the trained group ($M = 45.046, SD = 17.324 \mid t(69) = (2.905, p < .01$, see Figure 3.9). Contrary to our expectations, **the trained group reported significantly higher workload and significantly lower usability than the untrained group did** after completing knowledge exchange tasks with CLAICA (RQ2.1).

Last, we explored whether previous training influences user experience (UX) as

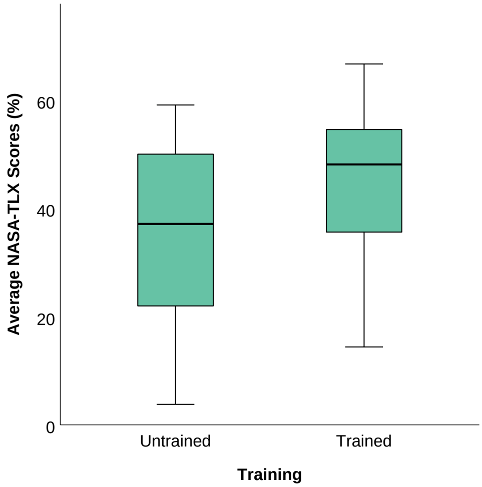

Figure 3.8: NASA-TLX score versus training

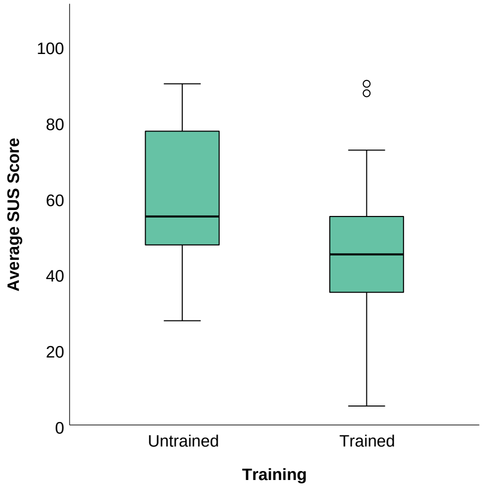

Figure 3.9: SUS score versus training

reported by our participants using the six dimensions of the UEQ. A series of independent-samples t-tests revealed significant differences between the two groups reflected in the average scores of **attractiveness** (untrained: $M = .392, SD = .854$ vs. trained: $M = -.133, SD = .848 \mid t(69) = (2.905, p < .05)$, **perspicuity** (untrained: $M = .853, SD = 1.284$ vs. trained: $M = -.133, SD = .848 \mid t(69) = (2.905, p < .01)$, **efficiency** (untrained: $M = .7647, SD = 1.191$ vs. trained: $M = .046, SD = .874 \mid t(69) = (2.905, p < .01)$, and **dependability** (untrained: $M = .471, SD = .824$ vs. trained: $M = .130, SD = .867 \mid t(69) = (1.430, p = .157)$, but not in the average scores of

stimulation (untrained: $M = .088, SD = .824$ vs. trained: $M = -.301, SD = .822$ | $t(69) = (1.702, p = .093)$ and **novelty** (untrained: $M = .1765, SD = .822$ vs. trained: $M = -.181, SD = .972$ | $t(69) = (1.395, p = .167)$. For an overview, see Figure 3.10. These findings indicate that **prior training in using CLAICA impacted negatively UX notions such as perceived attractiveness, perspicuity, efficiency, and dependability, but did not affect perceived stimulation and novelty (RQ2.1)**.

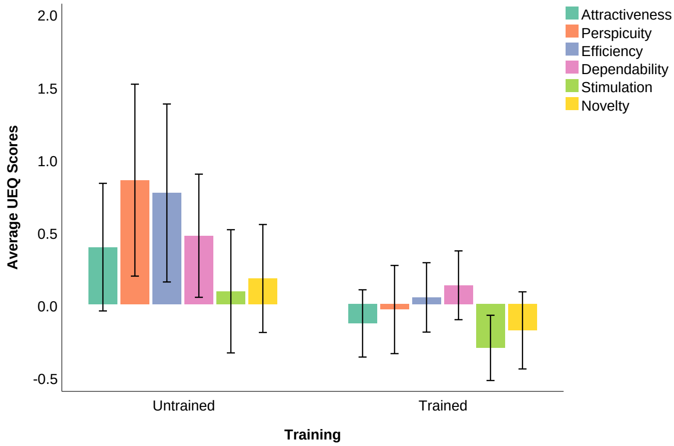

Figure 3.10: UEQ scores versus training

3.5.2. EFFECTS OF CONTEXT EXPERTISE

Next, we investigate if and how context expertise—being a **operator** vs. being a **layman**—affects task performance, perceived workload, usability, and UX when using a cognitive assistant such as CLAICA. Two non-parametric Mann-Whitney U test revealed significant differences in the median overall **task completion times** (operator: $Mdn = 502.985$ vs. layman: $Mdn = 380.775$ | $U = 261, p < .05$) and in the median overall **task completion rates** (operator: $Mdn = 7.5$ vs. layman: $Mdn = 7$ | $U = 333, p = .211$) between the two context-expertise groups (see Figure 3.11). Interestingly, **laymen performed better than operators in knowledge exchange tasks when using CLAICA (3.4.1)**. A follow-up non-parametric Mann-Whitney U test did not reveal significant differences in the median NASA-TLX scores between operator ($Mdn = 37.88$) and layman ($Mdn = 45.45$) participants ($U = 334.5, p = .236$).

However, an independent-samples t-test revealed a significant difference in the average SUS scores between operator ($M = 74.167, SD = 14.706$) and layman ($M = 48.521, SD = 18.896$) participants ($t(81) = -4.470, p < .001$)(see Figure 3.12). The mean SUS score from the operator and layman groups is equivalent to a "good" and "poor"/"ok" grade, respectively [161] or the 70th and 10th percentile, respectively [162]. Note that these equivalent scores predate widespread user testing of chatbots and may be unreliable [163]. These findings suggest that **laymen did not experience CLAICA as more cognitive demanding than operators in completing knowledge**

3

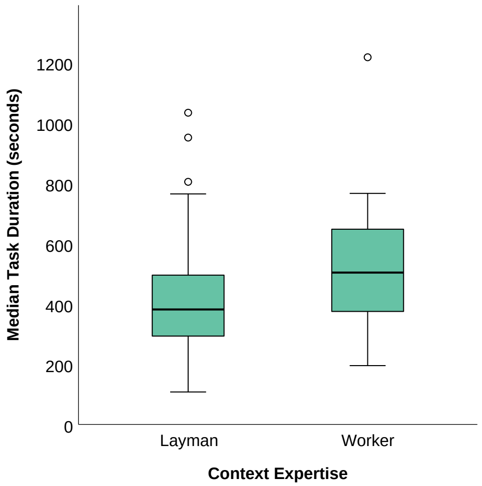

Figure 3.11: Task duration versus context expertise

exchange tasks. However, operators rated CLAICA significantly higher than laymen in terms of usability. (RQ2.2).

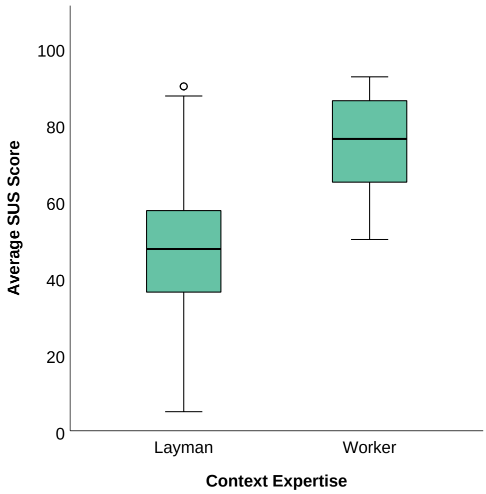

Figure 3.12: SUS score versus context expertise

Finally, we investigated whether and how context expertise has an impact on UX as reported by our participants using the six dimensions of the UEQ. A series of independent-samples t-tests revealed significant differences between the two groups reflected in the average scores of **attractiveness** (operator: $M = 1.305, SD = .887$ vs. layman: $M = -.007, SD = .872 \mid t(81) = -4.808, p < .001$), **perspicuity** (operator: $M = 1.458, SD = .851$ vs. layman: $M = .176, SD = 1.207 \mid t(81) = -3.525, p < .001$),

efficiency (operator: $M = 1.416, SD = .807$ vs. layman: $M = .218, SD = .999$ | $t(81) = -3.937, p < .001$), **dependability** (operator: $M = 1.208, SD = .744$ vs. layman: $M = .211, SD = .863$ | $t(81) = -3.765, p < .001$), **stimulation** (operator: $M = 1.333, SD = .807$ vs. layman: $M = -.207, SD = .833$ | $t(81) = -5.949, p < .001$), and **novelty** (operator: $M = .854, SD = 1.245$ vs. layman: $M = -.095, SD = .926$ | $t(81) = -3.117, p < .05$). For an overview, see Figure 3.13. **These findings indicate that context expertise plays a substantial role in UX when using a cognitive assistant such as CLAICA (RQ2.2).**

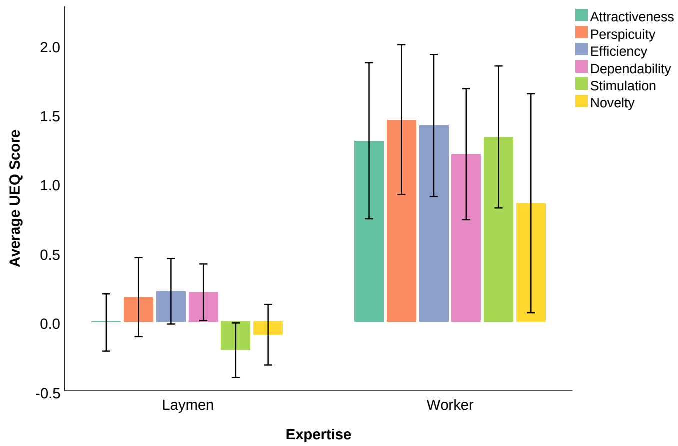

Figure 3.13: UEQ aspects versus context expertise

3.5.3. EFFECTS OF INTERACTION MODALITY

Then, we investigated if and how the interaction modality participants used (smartphone vs. laptop) had any impact on user performance, perceived workload, user experience and perceived usability. The aim here is to compare between a popular interaction modality (smartphone) and an established one (laptop) in how they influence information exchange with cognitive assistants such as CLAICA. From the outset, two non-parametric Mann-Whitney U tests displayed no significant differences in the median overall **task completion times** (smartphone: $Mdn = 385.981$ vs. laptop: $Mdn = 379.522$ $/mid$ $U = 765, p = .827$) and in the median overall **task completion rates** (smartphone: $Mdn = 7$ vs. laptop: $Mdn = 7$ | $U = 736.500, p = .296$). These findings indicate the **interaction modality (smarthphone vs. laptop) had no effect on task performance when executing knowledge exchange tasks with a cognitive assistant such as CLAICA (3.4.1).** In the same guise, two follow-up independent samples t-tests did not unveil any significant differences in the average NASA-TLX (smartphone: $M = 41.646, SD = 17.420$ vs. laptop: $M = 43.859, SD = 12.663$ | $t(61.410) = -.643, p = .523$) and in the average SUS (smartphone: $M = 53.333, SD = 25.663$ vs. laptop: $M = 51.383, SD = 15.462$ | $t(53.994) = .403, p = .688$) scores between the two interaction modality groups. These findings suggest that **interaction modality (smartphone vs. laptop) does not bear a substantial impact on perceived workload and reported usability when using a cognitive assistant such as CLAICA to perform knowledge exchange tasks (RQ2.3).**

Last but not least, we inquired into if and how interaction modality has influences UX as reported by our participants using the six dimensions of the UEQ. However, a series of independent-samples t-tests and Mann-Whitney U tests displayed no significant differences between the two groups as reflected in the average and median scores of **attractiveness** (smartphone: $M = .153, SD = 1.121$ vs. laptop: $M = .206, SD = .880 \mid t(81) = -.241, p = .810$), **perspicuity** (smartphone: $M = .361, SD = 1.495$ vs. laptop: $M = .362, SD = 1.029 \mid t(59.110) = -.002, p = .998$), **efficiency** (smartphone: $M = .542, SD = 1.157$ vs. laptop: $M = .2766, SD = .973 \mid t(81) = 1.133, p = .260$), **dependability** (smartphone: $M = .333, SD = 1.023$ vs. laptop: $M = .372, SD = .832 \mid t(81) = -.192, p = .849$), **stimulation** (smartphone: $Mdn = 0.250$ vs. laptop: $Mdn = 0.0 \mid U = 821, p = .818$), and **novelty** (smartphone: $M = -.208, SD = .992$ vs. laptop: $M = .234, SD = 1.021 \mid t(81) = -1.981, p = .051$). These findings show that **interaction modality (smartphone vs. laptop) has no effect on UX when using a cognitive assistant such as CLAICA to perform knowledge exchange tasks (RQ2.3).**

3.5.4. Qualitative Insights

We conducted a content analysis to create a structured representation of the comments received from our participants. At the first level, two coding categories were created: conversational AI perspective and UI perspective. At the second level, for the conversational AI perspective, text inputs were categorized according to the main technical components of the assistant which was built using the Rasa framework; such as NLU component, dialog management component, and external data bases. Two other categories include the assistant's functionalities and UX directly bound to the assistant. Comments related to the UI perspective were categorized into second-level aspects such as interaction modes, input modes, accessibility, general UX bound with the UI interactions, and message visualization. The comments were then further refined through emergent coding, as shown in Figure 3.14 and 3.15.

We identified 34/148 comments on the limited abilities of the NLU in understanding divergent phrasing resulting in conversational breakdown (NLU fallback). Four participants ($N = 4$) found the applied repair strategy and restart option to recover from the breakdown ineffective. Five participants ($N = 5$) wished they could talk to the assistant by using keywords instead of complete sentences. Participants expressed interest in the prospect of using CLAICA; One operator excitably asked "*when will it be able available in the app store*"(W13) and "*the ease of use and quick access to information are great*" (W13). Seven participants ($N = 7$) voiced concerns regarding safeguarding the quality of acquired knowledge and how to select the most appropriate recommendations as the knowledge base grows. Eleven participants ($N = 11$) stated that CLAICA might not be realistic or complex enough to be adopted in the manufacturing use case. One participant ($N = 1$) said "*seemed to work well for the prescribed tasks on the whole. Hard to relate this to real 'life on the floor', where*

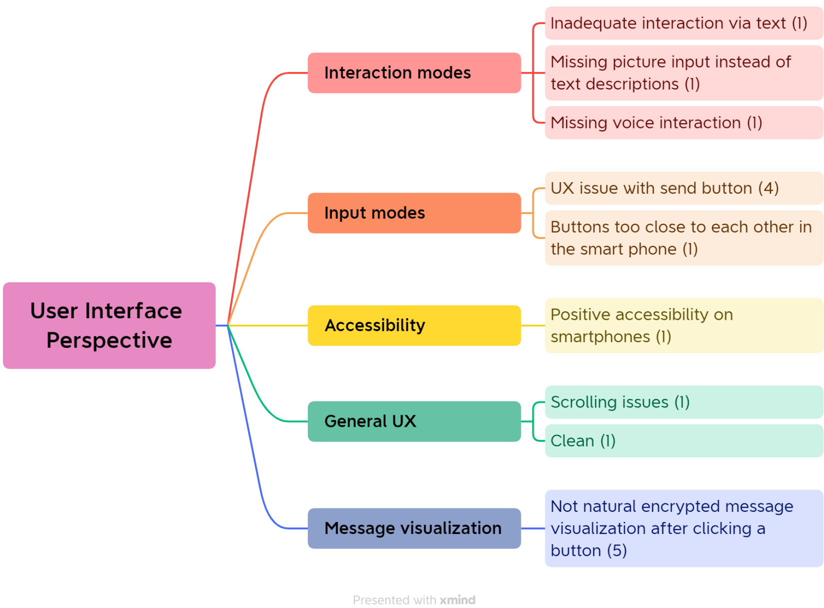

Figure 3.14: Results of the content analysis performed on participants' comments about the user interface (UI) perspective

presumably numerous kinds of errors (not all related to the machine) might occur." (L103) Another participant stated that "*operator should fix the issue or contact a supervisor if they cannot fix it, before they talk with chatbot*" (L13).

3.6. Discussion

One of CLAICA's major strengths is its ability to continuously learn from (expert) users; however, there are several hurdles that CLAICA must overcome to be effective. Many of these were raised by the participants; namely, (1) that its knowledge base was currently too limited, (2) how to ensure (long-term) accuracy, and (3) how to prevent information overload. As CLAICA can learn, its knowledge base does not need to be (quasi)complete before introduction; however, it would probably promote user acceptance if it could provide some level of assistance straight away. Therefore, we populated it with information about 100 previous problems that were solved by operators and a few product settings. We considered this sufficient for this user study; however, we aim to collect more before deploying it on a production line. We have already introduced a rating and approval mechanism that allows continuous feedback and flagging of inaccurate information. We may also consider prompting expert users to review information that CLAICA suspects is outdated (e.g., by its age or low rating). To tackle information overload, we rank the information in the knowledge base by several factors (e.g., rating and age) and only display the top

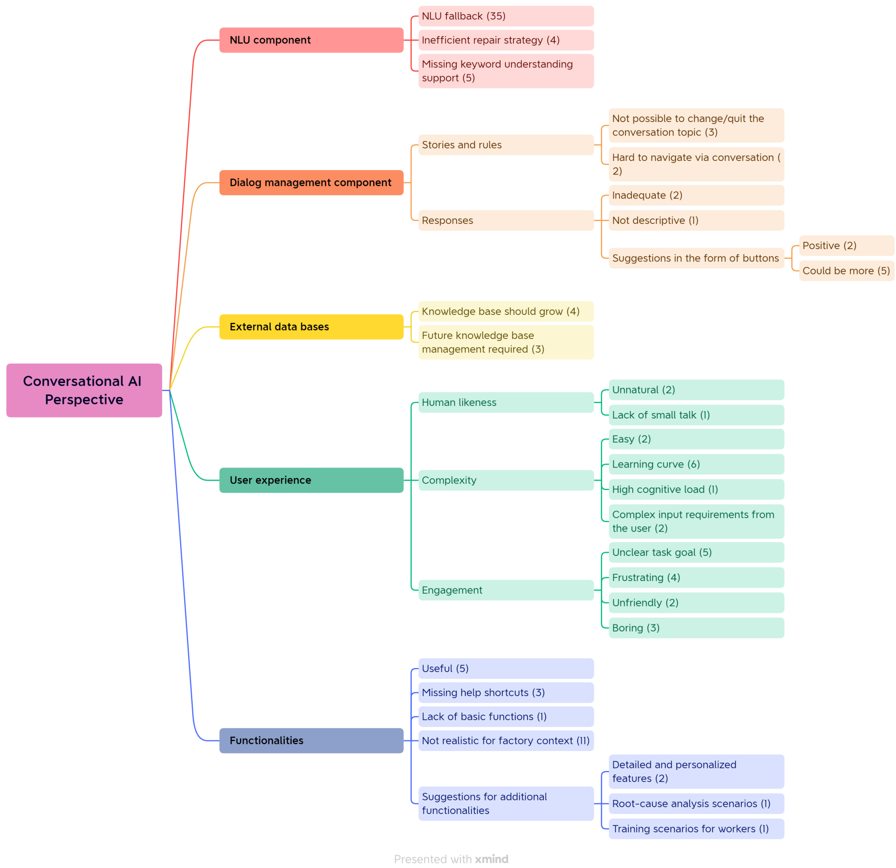

Figure 3.15: Results of content analysis performed on participants' comments about the conversational AI perspective

three to the users. However, we plan to use similarity algorithms on the knowledge graph to make this ranking more intelligent, for example, by suggesting knowledge about similar problems. Based on our experience designing and evaluating CLAICA, we created a set of design guidelines for future cognitive assistants.

3.6.1. Deprioritize Training when Users are Technologically Adept

The main advantage of conversational agents lies in that user interaction is greatly facilitated by the use of natural language. That said, we theorized that to perform knowledge exchange tasks with a cognitive assistant such as CLAICA, prior training should be necessary to fully realize its potential. However, our results showed that the trained group found the workload (NASA-TLX) to be higher, usability (SUS) lower, and attractiveness, perspicuity, and efficiency UEQ also lower. All other measures were found to be not significantly different between the trained and untrained groups. This is remarkable, as providing training for the assistant did not result in higher overall task completion rates or faster task completion time. Perhaps users that received training tried to remember what they had seen in the training session. In turn, this may have resulted in experiencing a higher workload, lower perceived usability and worse UX than the untrained participants did. Alternatively, the training may have set their expectations too high and they were disappointed when CLAICA failed to understand their utterances, resulting in inefficient and unclear exchanges. Future work is necessary to deep-dive and reliably pinpoint the cause of this effect.

3.6.2. Test with Diverse User Groups to Account for Vested Interest

Social desirability is a type of response bias and describes the tendency of people to present themselves in a generally favorable manner [164]. Typically, one would assume that the social desirability bias would manifest equally in the operator and layman participant groups, thus canceling itself out. operators rated CLAICA's usability (SUS) and user experience (UEQ) significantly higher than laymen. This result could be explained by the following factors: First, operators may (unconsciously) compare CLAICA with systems in their workplace that have poor usability and UX; second, operators are positively biased toward CLAICA as it was designed to help them and their colleagues (vested interest); and third, older operators might be more likely to be impressed by new technology. In fact, we found a significant correlation between age and SUS scores and there was a significant difference between the median age of the operators and laymen group. This may (partially) explain the difference in SUS scores between operators and laymen. Although there were no significant differences in perceived workload (NASA-TLX) scores or overall task completion rates, operators took significantly longer to complete the tasks. Although the operators group had experience with factory production lines, this clearly did not help them interact more efficiently with CLAICA. In fact, it appears to be detrimental. Maybe the operators took longer because they were on average older and had worse digital competencies. These findings have implications for studies that use laymen to test systems for a specific group of end-users. Namely, that context expertise did not help the user interact more efficiently with the assistant, help them complete more tasks, or reduce workload. However, context expertise positively affected subjective usability and UX.

3.6.3. Intent Shortcuts Override Interaction Modality Types

The interaction modality (smartphone or laptop) appeared to not significantly affect any of the measures. This is unexpected, as the participants who used a laptop to interact with CLAICA could probably type faster [157], had a larger screen at their disposal, and in theory should have completed the knowledge exchange tasks faster. Perhaps the speed-typing advantage on laptops was counteracted by the availability of auto correct on the participants' smartphones and, perhaps more importantly, by

CLAICA's use of buttons/shortcuts. Alternatively, it is also possible the time required for the typing did not constitute a significant part of the duration of the task. Considering that there appears to be no benefit to using a laptop over a smart, we suggest interacting with cognitive assistants like CLAICA on a smartphone, as many operators already carry one around in their pocket, and it can be used at whatever location is relevant to the required cognitive assistance or knowledge sharing.

3.6.4. Salvage Conversational Breakdowns and Support Single-Word Input

We grouped the comments of our participant into the following two categories: Conversational AI and UI. The effectiveness of CLAICA's NLU was the most prominent subtopic, for example, how well CLAICA could understand user intents and handle fallback. This is not surprising considering the challenges AI has in understanding natural language and the limited training examples (20-50 examples per intent) we used to train CLAICA. Regarding the divergent phrasing of user utterances, additional training data (e.g., from this user study) can be used to improve its accuracy. As it is important to test frequently and early with users, we cannot avoid NLU breakdowns; however, having a robust fall-back mechanism (e.g., offering the top three intent predictions) can soften the impact on UX.

Five participants, including factory operators who wanted to have short interactions, suggested providing a better NLU for keywords (that is, utter a single word to indicate intent). However, training an NLU model to classify intents purely based on a keyword may reduce its accuracy if the same keyword could be associated with several intents. Nevertheless, using keywords for frequently used features may work well, as we recognize that this is natural for humans and can save time.

3.6.5. Use Suggested Responses to Streamline Interaction

In contexts where users frequently use the same assistant (e.g., at a workplace or home), it is more reasonable to expect users to learn specific commands to invoke a feature. Indeed, several participants also mentioned that there was a learning curve in interacting with CLAICA as fully natural language did not always work. Nevertheless, many participants were positive about the interactive efficiency they could achieve once they learnt how to make themselves understood by CLAICA, for example, through the use of its shortcut buttons or specific phrasing. These buttons enabled faster responses and gave users hints as to what response was expected from them. Considering their success, we will continue to implement them and explore how to make the suggestions more intelligent and informative.

3.6.6. Ethical Considerations

The use of a system such as CLAICA raises several ethical concerns. First, we do not know what the long-term effects of using CLAICA may be. CLAICA can help share knowledge among operators; however, they can become overly dependent on its recommendations. Perhaps CLAICA could ask open questions to operators to invoke critical thinking. Second, the real-time data that CLAICA uses can be used by management to infringe on operators' privacy rights. Therefore, it is important that managers (or other operators) cannot use CLAICA to track people's performance or interactions without their knowledge. As such, guidelines and policies to protect and respect user privacy and rights, such as data management plans, are fundamental to Industry 5.0. Finally, it is not clear who owns the knowledge that operators share with CLAICA. One could claim that the knowledge acquired from the operators and formalized by CLAICA, becomes company property. In turn, by sharing their knowledge, operators effectively reduce their (perceived) value. Therefore, it is important that CLAICA or their employer employs reciprocal strategies. For example, the factory collaborating with the development of CLAICA has introduced a monetary reward system for operators who shared the most high-quality knowledge over a month period.

3.6.7. Limitations

This study focused on the task performance, usability, and user experience through several interactions with CLAICA; however, participants did not have to act on the exchanged information or share their own knowledge. Furthermore, the interactions were performed in the lab ("in vitro") as opposed to an actual production line. Therefore, the results reflect a part of CLAICA's UX. Assessing the UX of CLAICA "in vivo" will be the focus of future studies. To keep the duration of the user study to a reasonable duration (about an hour), we informed test participants that they could begin completing the survey after spending ten minutes trying to complete the assigned tasks. Although the vast majority of participants completed the tasks in less than ten minutes, several participants took longer, including multiple factory operators (see Figure 3.11). Considering that this affected all groups and that the sample size for the factory was relatively small ($N = 12$), we decided not to remove these data points. We found a correlation between age and SUS. This may partially explain the difference in SUS for the context expertise dimension; however, we believe the difference in age between the groups to be indicative of the reality in factories. That is, many operators are >50 years old and have poorer digital competencies than many of the younger laymen we recruited for the study. Finally, we conducted the study in six user trials with two different moderators and en masse. First, conducting the study in a group setting may affect the results, as participants may feel pressure to finish faster than their peers or learn from each other. We tried to minimize this effect by asking participants not to converse with each other, anonymizing the results, and collecting the task performance measures automatically so that participants did not have to indicate when they were done. Second, differences in how the sessions were conducted, such as the wording or demeanor of the moderators, could also affect the results. We believe we minimized the effects sufficiently by following a script for the instructions and using the same slides, task instructions, and video material for the user trials.

3.7. Conclusion

We presented CLAICA, a continuously learning AI cognitive assistant that supports agile manufacturing operators by exchanging knowledge and providing quick access to instructional material. CLAICA is the product of a co-design process with factory operators and the focus of a user study with 83 participants. Our findings contribute to a deeper understanding of how prior training, context expertise, and interaction modality affect the user experience of cognitive assistants. Drawing on our findings, we elicit design and evaluation guidelines for cognitive assistants that support knowledge exchange in cognitively demanding tasks and challenging environments (e.g., an agile production line).

In this chapter, we evaluated CLAICA alongside a simulated environment and text-only interaction, observing the shortcomings of traditional natural language processing, such as frequent misunderstandings and the inefficiencies of text-only interaction. To mitigate some of these problems, we will explore the potential of integrating large language models (LLM) in cognitive assistants in Chapter 4.

4

INTEGRATING LARGE LANGUAGE MODELS IN COGNITIVE ASSISTANTS

To overcome the understanding limitations of traditional intent-based conversational AI, such as the cognitive assistant system presented in Chapter 2, we investigated how to take advantage of recent advances in natural language processing (NLP), namely large language models (LLM). This chapter introduces a Large Language Model-based system designed to retrieve information from the extensive knowledge contained in factory documentation and knowledge shared by expert operators. The system aims to efficiently answer queries from operators and facilitate the sharing of new knowledge. We conducted a user study at a factory to assess its potential impact and adoption, eliciting several perceived benefits, namely, enabling quicker information retrieval and more efficient resolution of issues. However, the study also highlighted a preference for learning from a human expert when such an option is available. Furthermore, we benchmarked several proprietary and open-sourced Large Language Models (LLMs) for this system. GPT-4, representing a state-of-the-art LLM at the time of writing, consistently outperformed its counterparts, with open-source models trailing closely, presenting an attractive option given their data privacy and customization benefits. In summary, this work offers insights and a system design for factories considering using LLM tools for knowledge management in factories.

The content of this chapter is derived from the work published in S. Kernan Freire, C. Wang, M. Foosherian, S. Wellsandt, S. Ruiz-Arenas, and E. Niforatos. “Knowledge sharing in manufacturing using LLM-powered tools: user study and model benchmarking”. In: *Frontiers in Artificial Intelligence* 7 (2024). ISSN: 2624-8212. DOI: 10.3389/frai.2024.1293084.

4

4.1. Introduction

Human-centric manufacturing seeks to harmonize the strengths of humans and machines, aiming to enhance creativity, human well-being, problem-solving abilities, and overall productivity within factories [165–167]. Despite these advancements, a significant challenge persists in effectively managing and utilizing the vast knowledge generated within these manufacturing environments, such as issue reports and machine documentation [168]. This knowledge is crucial for optimizing operations, yet it remains largely untapped due to the difficulties in processing and interpreting the disconnected, sometimes unstructured, technical information it contains [169].

Traditionally, leveraging this knowledge has been cumbersome, with operators choosing to use personal smartphones over official procedures [170] and AI unable to handle the complexity of the data [16]. However, recent Large Language Models (LLMs) like GPT-4 show promise in addressing these challenges. LLMs can effectively interpret, summarize, and retrieve information from vast datasets [99] while concurrently aiding the capture of new knowledge [171]. These capabilities could significantly support operators in knowledge-intensive tasks, making it easier to access relevant information, share new knowledge, and make informed decisions rapidly.

In response to the challenge that factories face in sharing knowledge among operators, we developed a CA to leverage factory documents and issue analysis reports to answer operators' queries. Besides answering queries on existing documentation, the tool facilitates the analysis and reporting of new issue solutions. Despite the promising capabilities of LLMs, their application in manufacturing is not straightforward. The specific, dynamic knowledge required in this domain poses unique challenges [172]. For instance, a foundational LLM may have limited utility in a factory setting without significant customization, such as fine-tuning or incorporating specific context information into its prompts [94]. Additionally, the practical and socio-technical risks and challenges of deploying LLMs in such environments remain largely unexplored—factors that are key to successfully implementing human-centered AI [101]. Concerns include the accuracy of the information provided, the potential for "hallucinated" answers [173], and the need for systems that can adapt to the highly specialized and evolving knowledge base of a specific manufacturing setting [172]. This CA demonstrates the feasibility of using LLMs to enhance knowledge management in manufacturing settings. To understand its effectiveness and potential, we conducted a user study in a factory environment, evaluating the system's usability, user perceptions, adoption, and impact on factory operations.

This work also addresses the lack of specific benchmarks for evaluating LLMs in manufacturing. We benchmarked several LLMs, including both closed and open-source models, recognizing that the standard benchmarks[1] primarily focus on general knowledge and reasoning. As such, they may not adequately reflect the challenges of understanding manufacturing-specific terminology and concepts. This benchmarking focused on their ability to utilize factory-specific documents and

[1] https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard—last accessed November 20, 2024

unstructured issue reports to provide factual and complete answers to operators' queries.

In summary, this chapter explores the intersection of LLM technology and knowledge management in the manufacturing industry. It discusses the development and evaluation of an LLM-powered tool designed for manufacturing environments and presents findings to inform future efforts to use this technology in similar settings. The contributions of this Chapter can be summarized as follows:

1. We present the system design of an LLM-powered cognitive assistant to acquire and share knowledge with factory operators.
2. We provide a deeper understanding of the opportunities, challenges, and risks associated with using LLM-powered cognitive assistants by factory operators.
3. We present a benchmark of LLMs for retrieval augmented generation on manufacturing content.

4.2. BACKGROUND

In this section, we address the topic of industry 5.0, LLM-powered tools for knowledge management, benchmarking LLMs, and the research questions informing this work.

4.2.1. HUMAN-CENTERED MANUFACTURING

Industry 5.0, the latest phase of industrial development, places human beings at the forefront of manufacturing processes, emphasizing their skills, creativity, and problem-solving abilities [165]. Human-centered manufacturing in Industry 5.0 focuses on providing a work environment that nurtures individuals' creativity and problem-solving capabilities [174]. It encourages workers to think critically, innovate, and continuously learn. With machines handling repetitive and mundane tasks, human workers can dedicate their time and energy to more complex and intellectually stimulating activities. This shift could enhance job satisfaction and promote personal and professional growth, as workers could acquire new skills and engage in higher-level decision-making [165, 175]. Emphasis on human-machine collaboration and the continuous emergence and refinement of technology increases the need for adequate human-computer interaction [176]. One of the approaches to address this topic is using conversational AI to assist humans in manufacturing [177].

4.2.2. LLM-POWERED KNOWLEDGE MANAGEMENT TOOLS

Training Large Language Models (LLMs) on numerous, diverse texts results in the embedding of extensive knowledge [88]. LLMs can also adeptly interpret complex information [87], general reasoning [89], and aiding knowledge-intensive decision-making. Consequently, researchers have been exploring applying LLM-powered tools in domain-specific tasks [178–180].

Despite their potential benefits, the responses generated by LLMs may have two potential issues: (1) outdated information originating from the model's training date, and (2) inaccuracies in factual representation, known as "hallucination" [88, 95]. To address these challenges and leverage the capabilities of LLMs in domain-specific knowledge-intensive tasks, several techniques can be used, such as chain-of-thought [181], few-shot prompting [97, 98], and retrieval augmented generation [99].

Using few-shot prompting to retrieve information across diverse topics, Semnani *et al.* [182] introduced an open-domain LLM-powered chatbot called WikiChat. WikiChat utilizes a 7-stage pipeline of few-shot prompted LLM that suggests facts verified against Wikipedia, retrieves additional up-to-date information, and generates coherent responses. They used a hybrid human-and-LLM method to evaluate the chatbot on different topics for factuality, alignment with real-worth truths and verifiable facts, and conversationality. This compound metric scores how informational, natural, non-repetitive, and temporally correct the response is. Their solution significantly outperforms GPT-3.5 in factuality, with an average improvement of 24.4% while staying on par in conversationality. Others have explored the capabilities of LLMs in domain-specific tasks such as extracting structured data from unstructured healthcare texts [183], providing medical advice [184], simplifying radiology reports [185], Legal Judgement Prediction from multilingual legal documents [186], and scientific writing [187].

Researchers are cautiously investigating the applications of LLMs in manufacturing. Xia *et al.* [93] demonstrated how using in-context learning and injecting task-specific knowledge into an LLM can streamline intelligent planning and control of production processes. Mercedes-Benz [92] used ChatGPT to enhance quality management and process optimization in vehicle production. Wang *et al.* [94] conducted a systematic test of ChatGPT's responses to 100 questions from course materials and industrial documents. They used a zero-shot method and examined the responses' correctness, relevance, clarity, and comparability. Their results suggested areas for improvement, including low scores when responding to critical analysis questions, occasional non-factual or out-of-manufacturing scope responses, and dependency on query quality. Although Wang *et al.* [94] provides a comprehensive review of ChatGPT's abilities to answer questions related to manufacturing; it did not include the injection of task-specific knowledge into the prompts.

To improve the performance of an LLM for domain-specific tasks, relevant context information can be automatically injected along with a question prompt. This technique, known as Retrieval Augmented Generation (RAG), involves searching a corpus for information relevant to the user's query and inserting it into a query template before sending it to the LLM [99]. Using RAG also enables further transparency and explainability of the LLM's response. Namely, users can check the referenced documents to verify the LLM's response. Factories will likely have a large corpus of knowledge available in natural language, such as standard work instructions or machine manuals. Furthermore, factory workers continually add to the pool of available knowledge through (issue) reports. Until recently, these reports were considered unusable by AI natural language processing due to quality issues such as poorly structured text, inconsistent terminology, or incompleteness [16]. However, the leap in natural language understanding that LLMs, such as ChatGPT, have brought about can overcome these issues.

4.2.3. EVALUATING LARGE LANGUAGE MODELS

Large Language Model evaluation requires the definition of evaluation criteria, metrics, and datasets associated with the system's main tasks. There are two types of LLM evaluations: intrinsic and extrinsic evaluation. Intrinsic evaluation focuses on the internal properties of a Language Model [188]. It means the patterns and language structures learned during the pre-training phase. Extrinsic evaluation focuses on the model's performance in downstream tasks, i.e., in the execution of specific tasks that make use of the linguistic knowledge gained upstream, like code completion [189]. Despite extrinsic evaluation being computationally expensive, only conducting intrinsic evaluation is not comprehensive, as it only tests the LLMs capability for memorization [190]. We will focus on extrinsic evaluation as we are primarily interested in the performance of LLM-based tools for specific real-world tasks.

Extrinsic evaluation implies assessing the systems's performance in tasks such as question answering, translation, reading comprehension, and text classification, among others [191]. Existing benchmarks such as HellaSwag [192], TriviaQA [193], and MMLU [194], among others, are widely reported in the literature for comparing language models. Likewise, domain-specific Benchmarks for tasks such as medical [195], fairness evaluation [180], finance [196], robot policies [197], and 3D printing code generation [198] can also be found. Experts also evaluate the performance of large-language models (LLMs) in specific downstream tasks, such as using physicians to evaluate the output of medical specific LLMs [195].

LLM benchmarks range from specific downstream tasks to general language tasks. However, to our knowledge, LLM models have not been benchmarked for answering questions in the manufacturing domain based on context material, a technique known as Retrieval Augmented Generation [99]. Material such as machine documentation, standard work instructions, or issue reports will contain domain jargon and technical information that LLMs may struggle to process. Furthermore, the text in an issue report may pose additional challenges due to abbreviations, poor grammar, and formatting [16]. Therefore, as part of this work, we benchmarked several LLM models on their ability to answer questions based on factory manuals and unstructured issue reports. Furthermore, we conducted a user study with factory operators and managers to assess the potential benefits, risks, and challenges. We defined the following dissertation research question, "**What are the implications of using LLM-based conversational AI for knowledge sharing among factory operators?**" (RQ3), which can be broken down into the following research questions:

RQ3.1 *What are the perceived benefits, challenges, and risks of using Large Language Models for information retrieval and knowledge sharing for factory operators?*

RQ3.2 *How do Large Language Models compare in performance when answering factory operators' queries based on factory documentation and unstructured issue reports?* We consider performance as the factuality, completeness, hallucinations, and conciseness of the generated response.

4.3. LLM-based System for Information Retrieval and Knowledge Sharing in Factories

We built a fully functional system to assess the potential of using LLMs for information retrieval and knowledge sharing for factory operators. Benefiting from LLMs' in-context learning capabilities, we use this to supply an LLM with information in the form of factory manuals, and issue reports relevant to the user's question, a technique known as Retrieval Augmented Generation (RAG) [99], see Figure 4.1[2]. As noted by [89], training LLMs using a prompt packed with query-related information can yield substantial performance enhancement. Users can ask questions in the chat box by typing or using voice input. The response is displayed at the top of the page, and the document chunks used for the answer can be checked at the bottom (see Figure 4.2).

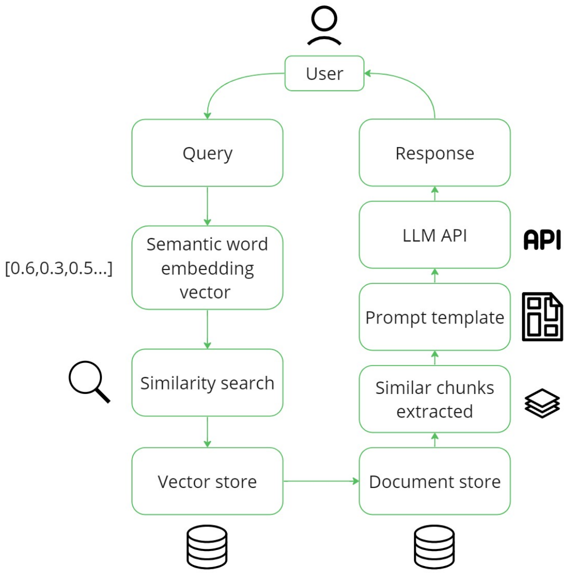

Figure 4.1: The steps of Retrieval Augmented Generation (RAG) from user query to response

Tool Dependencies

The tool was constructed utilizing two innovative technologies—Gradio [199] and LlamaIndex[3]. Gradio, a tool developed by Abid *et al.* [199], serves as the backbone for both our front and back ends. Primarily used to simplify the development and

[2]The system presented here is registered at the TU Delft patent department as eIDF 2023E00075 NL
[3]https://www.llamaindex.ai/—last accessed November 20, 2024

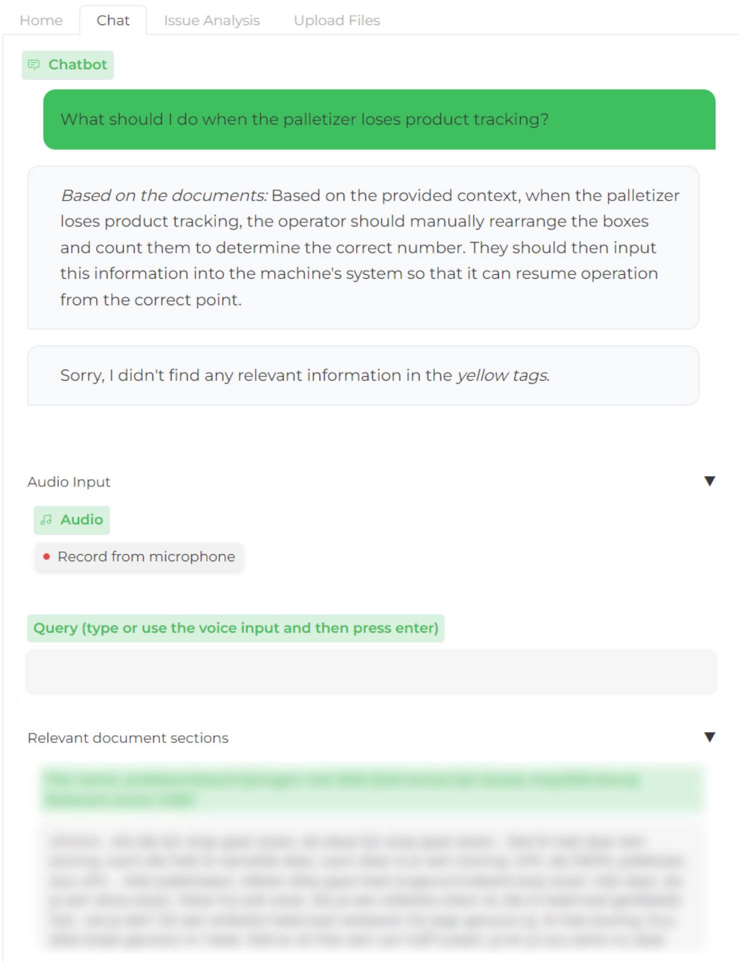

Figure 4.2: The main chat interface. The "relevant document sections" part is blurred for confidentiality as it shows the title of a company's document and its content.

distribution of machine learning applications, Gradio allows the quick creation of intuitive, user-friendly web interfaces for machine learning models.

Additionally, we use LlamaIndex, created by [200], for retrieving the context material in response to the user queries and handling the queries to the LLM. LlamaIndex, initially known as GPT Index, is a cutting-edge data framework designed for the efficient handling and accessibility of private or domain-specific data in LLMs applications.

Since the factory documents can be long, they may overflow the LLM's context window or result in unnecessary computational demand. To overcome this, we segment the materials into manageable chunks, each comprising approximately 400 tokens. This method effectively incorporates the materials into the LLM prompt without compromising the conversation flow. Following the segmentation, each document chunk is processed through LlamaIndex using the OpenAI Embedding API[4]. Utilizing the "text-embedding-ada-002" model, LlamaIndex transforms each chunk into a corresponding embedding vector. These resulting vectors are then securely stored, ready for future retrieval and use.

Knowledge Base Construction

Our experiment incorporates two distinct types of domain-specific data: factory manuals and shared knowledge from factory workers. Factory manuals outline information on machine operation, safety protocols, quality assurance, and more. These resources, provided by factory management teams, initialize the knowledge base for each specific factory. The materials come in various formats, including PDF, Word, and CSV files.

In addition to the factory manuals, we integrate issue analysis reports from factory workers. This information is gathered from the production line, utilizing the 5-why process, an iterative root-cause analysis technique [201] (see Figure 4.3). The 5-why technique probes into cause-and-effect relationships underlying specific problems by repeatedly asking "why?" until the root cause is revealed, typically by the fifth query. This process enables us to gather real-world issues encountered on production lines, which may not be covered in the factory manuals. Upon entering all required information, including one or more "whys", the operator presses "check", triggering a prompt to the LLM that performs a logical check of the entered information and checks for inconsistencies with previously reported information. The operator can revise the entered information and submit it as is. Then, the submitted report will be added to a queue for expert operators to check before it is added to the knowledge base.

Prompt Construction

To retrieve the document data relevant to specific user queries, we employ the same embedding model, "text-embedding-ada-002" to generate vector representations of these queries. By leveraging the similarity calculation algorithm provided by LlamaIndex, we can identify and retrieve the top-K most similar segmented document snippets related to the user query. This allows us to construct pertinent LLM queries. Once the snippets are retrieved, they are synthesized into the query template based on those used by LlamaIndex[5] as shown in Figure 4.4.

However, considering our data originates from two distinct sources—factory manuals and shared tactical knowledge—we have decided to segregate these into two separate LLM queries. This approach is designed to prevent potential user confusion from combining data from both sources into a single query.

[4] https://api.openai.com/v1/embeddings—last accessed November 20, 2024
[5] https://docs.llamaindex.ai—last accessed November 20, 2024

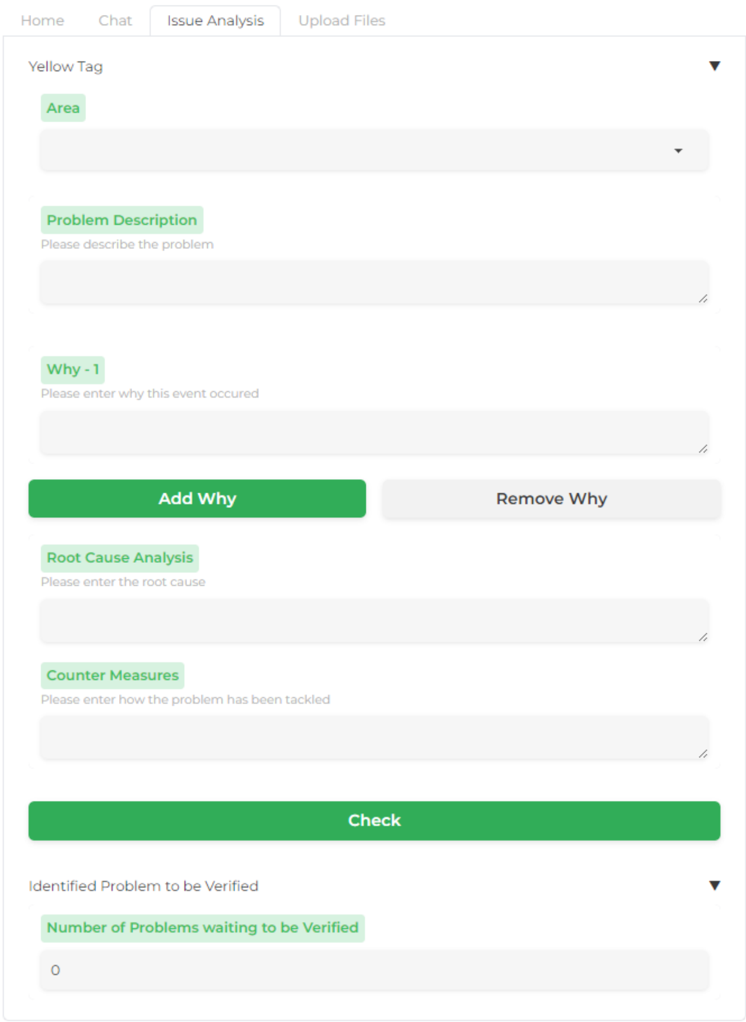

Figure 4.3: The issue analysis interface for describing a problem, root-cause and solution. When the operator clicks on 'Check', we use an LLM prompt to check the logic of the entered information and check for consistency with existing knowledge.

4.4. USER STUDY IN THE FIELD

We conducted a user study on the system to uncover perceived benefits, usability issues, risks, and barriers to adoption. The study comprised three tasks: to ask the system several questions as if they were operators, to fill in a "yellow tag" (issue analysis report) based on a recent issue and request a logical check, and finally, to

```
You are an assistant that assists detergent production line
operators with decision support and advice based on a knowledge
 base of standard operating procedures, single point lessons (
SPL), etc. We have provided context information below from
relevant documents and reports.

------------------------------------------------------------
                [Retrieved Document Snippets]
------------------------------------------------------------

Given this information, please answer the following question:
                        [Query]

If the provided context does not include relevant information
to answer the question, please do not respond.
```

Figure 4.4: Retrieval augmented generation prompt template.

upload new documents to the system. After each task, they were asked to provide feedback. Then, after completing all tasks, the participants were posed several open questions about the system's benefits, risks, and barriers to adoption. Finally, demographic information, such as age, gender, and role, was requested.

4.4.1. Participants

We recruited $N = 9$ participants from a detergent factory, of which $n = 4$ were managers (P1-4), and $n = 5$ were operators (P5-9). Of the nine participants, $n = 3$ were women, and $n = 6$ were men. Participant age was distributed over three brackets, namely $n = 2$ were 30-39, $n = 4$ were 40-49, and $n = 3$ were 50-59.

4.4.2. Qualitative Analysis

An inductive thematic analysis [202] of the answers to the open questions resulted in six themes discussed below.

- **Usability** The theme of usability underlines the system's ease of use and the need for clear instructions. Users mentioned the necessity for a "user-friendly" (P2) interface and highlighted the importance of having "more instructions and more details need to be loaded" (P1) to avoid confusion. This indicates a desire for intuitive navigation that could enable workers to use the system effectively without extensive training or referencing external help. The feedback suggests that the system already works well, as reflected in statements like "Easy-to-use system" (P3) and the system "works well" (P7).

- **Access to information** Users appreciated the "ease of having information at hand" (P1), facilitating immediate access to necessary documents. However, there is a clear call for improvements, such as the ability to "Include the possibility of opening IO, SPL, etc. in .pdf format for consultation" (P3). This theme is supported by requests for direct links to full documents, suggesting that while "the list of relevant documents from which information is taken is excellent" (P4), the ability to dive deeper into full documents would significantly enhance the user experience.

- **Efficiency** Users value the "greater speed in carrying out some small tasks" (P3). However, there are concerns about the system's efficiency when it does not have the answer, leading to "wasting time looking for a solution to a problem in case it is not reported in the system's history" (P3). Statements like "quick in responses" (P3) contrast with the need for questions to be "too specific to have a reliable answer" (P7), indicating tension between the desire for quick solutions and the system's limitations.

- **Adoption** Users highlight several factors affecting adopting the new system. It includes challenges such as "awareness and training of operators [might hinder adoption]" (P3) and the need for "acceptance by all employees" (P4), which indicates that the system's success is contingent on widespread user buy-in. The generational divide is also noted: "That older operators use it [on what may hinder adoption]" (P7) suggests that demographic factors may influence the acceptance of new technology.

- **Safety** Users express apprehension that "if the responses are not adequate, you risk safety" (P1), emphasizing the critical nature of reliable information in a factory setting. Moreover, the demand for updated and specific information underlines the importance of the system's content being current and detailed to maintain operational safety standards, as stated by P9: "If it is updated and specific, it can help me".

- **Traditional versus Novel** There is a noticeable preference for established practices among some users. For instance, "It's faster and easier to ask an expert colleague working near me rather than [the system]" (P8) captures the reliance on human expertise over the assistant system. This tension is further demonstrated by the sentiment that "Operators may benefit more from traditional information retrieval systems" (P9), suggesting a level of skepticism or comfort with the status quo that the new system needs to overcome.

4.5. LARGE LANGUAGE MODEL BENCHMARKING

In our benchmarking experiment, we evaluated various proprietary and open-source LLMs, including OpenAI's ChatGPT (GPT-3.5 and GPT-4 from July 20$^{\text{th}}$, 2023), Guanaco 65B and 35B variants [203] based on Meta's Llama (Large Language Model Meta AI) [204], Mixtral 8x7b [205], Llama 2 [206], and one of its derivatives, StableBeluga2 [207]. This selection represents the state-of-the-art closed-sourced models (e.g., GPT-4) and open-source models (e.g., Llama 2). We included the (outdated) Guanaco models to demonstrate the improvements in the open-source sphere from March 2023 to July 2023.

We used a web UI for LLMs[6] to load and test the Mixtral 8x7B, Guanaco models, and the StableBeluga2. The models were loaded on a pair of Nvidia A6000s with NVlink and a total Video Random Access Memory (VRAM) capacity of 96 GB. The 65B model was run in 8-bit mode to fit in the available VRAM. We used the llama-precise parameter preset and fixed zero seed for reproducibility. Llama 2 was evaluated using the demo on huggingface[7].

To rigorously assess the models, we prepared 20 questions of varying complexity based on two types of context material: formal operating material and informal issue reports. The model prompt was constructed using the above template (4.4). Ultimately, the difficulty of a question is a combination of the question's complexity and the clarity of the source material. Simple questions include retrieving a single piece of information clearly stated in the context material, for example, "At what temperature is relubrication necessary for the OKS 4220 grease?". Conversely, difficult questions require more reasoning or comprise multiple parts, for example, "What should I do if the central turntable is overloaded?" which has a nuanced answer dependent on several factors not clearly articulated in the context material.

In addition to measuring response length in words, we defined several metrics to qualitatively score the responses to measure their utility for factory operators, similar to prior work that evaluated a medical LLM [195] and general knowledge [182]. These metrics are factuality, completeness, and hallucinations, as defined below:

- **Factuality**: Responses align with the facts in the context material.
- **Completeness**: Responses contain all the information relevant to the question in the context material.
- **Hallucinations**: Response appears grammatically and semantically coherent but is not based on the context material.

The following scoring protocol is applied: one is awarded for a completely factual, complete, or hallucinated response. In contrast, a score of 0.5 is awarded for a slightly nonfactual, incomplete, or hallucinated response (e.g., the response includes four out of the five correct steps). Otherwise, a score of zero is awarded. Therefore, wrong answers are penalized heavily. If the model responds by saying it cannot answer the question and does not make any attempt to do so, it is scored zero for factuality and completeness, but no score is given for hallucination. As such, the final score for hallucination is calculated as follows: $\text{corrected score} = \frac{\text{score}}{20 - \text{number of unanswered questions}} \times 100$

As shown in Figure 4.5 and Table 4.1, GPT-4 outperforms other models regarding factuality, completeness, and lack of hallucinations but is closely followed by StableBeluga2 and GPT-3.5. The Guanaco models, based on Llama 1, perform significantly worse. The conciseness of the responses showed a similar pattern, except that StableBeluga2 produced the shortest answers (58 words), followed closely by Mixtral 8x7B (66 words) and GPT-4 (69 words).

[6] https://github.com/oobabooga/text-generation-webui/tree/main—last accessed November 20, 2024

[7] https://huggingface.co/meta-llama/Llama-2-70b-chat-hf—last accessed November 20, 2024

Table 4.1: Model Benchmarking Scores (out of 100) and Average Response Length

Model	**Factuality**	**Completeness**	**Hallucinations**	**Words**
GPT-4	97.5	95	0	69
StableBeluga2	95	92.5	7.5	58
Mixtral 8x7B	92.5	92.5	2.5	66
GPT-3.5	90	90	5	89
Llama 2	77.5	82.5	13	128
Guanaco 65B	55	39.5	65	131
Guanaco 33b	27.5	27.5	65.6	190

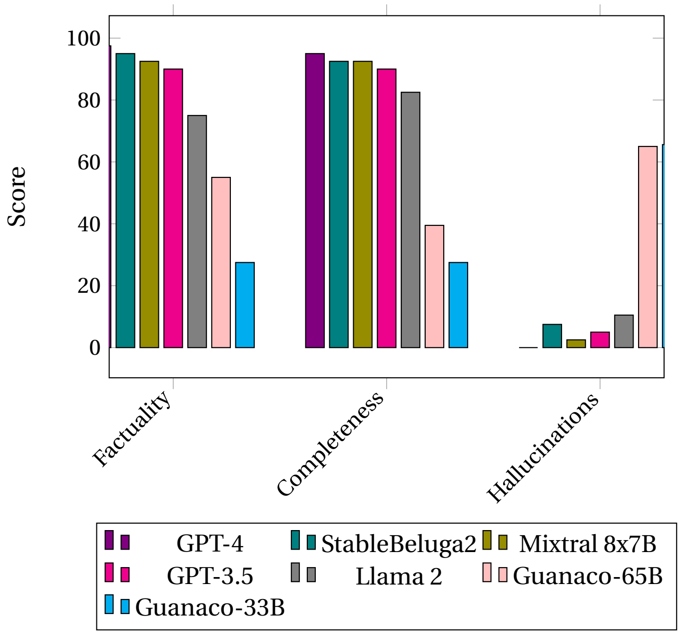

Figure 4.5: Benchmark of seven LLMs for generating answers based on factory materials.

4.6. Discussion

4.6.1. GPT-4 is the Best but Open-Source Models Follow Closely

GPT-4 performs best across all measures but is closely followed by StableBeluga2, Mixtral 8x7B, and GPT-3.5. Compared to GPT-4, the cost per input token for GPT-3.5 is significantly lower[8]. However, the higher costs of GPT-4 are partially counteracted by the shorter response length. GPT-3.5 (and Llama 2) tended to be

[8] https://openai.com/pricing#language-models—last accessed November 20, 2024

wordier and include additional details that were not directly requested, whereas GPT-4, StableBeluga2, and Mixtral 8x7B generated more concise responses.

The latest generation of open-source models, such as Mixtral 8x7B and Llama 2 variants, such as StableBeluga2, demonstrates a clear jump forward relative to their predecessors based on Llama-1, which were more prone to hallucinations and exhibited poorer reasoning abilities over the context material. While open-source models like StableBeluga2 and Mixtral 8x7B do not score as high as GPT-4, they ensure better data security, privacy, and customization if hosted locally. This can be a crucial consideration for companies with sensitive data or unique needs.

4.6.2. The Tool is Beneficial but Inferior to Human Experts

Users appreciate the system's functionality and see it as a tool for modernizing factory operations and speeding up operations. They are keen on improvements to be made for better user experience and utility, especially in the areas of content, feature enhancements, and user training. However, they express concerns about potential safety risks and the efficacy of information retrieval compared to consulting expert personnel. While these concerns are understandable, the tool was not designed to replace human-human interactions; instead, it can be used when no human experts are present or when they do not know or remember how to solve a specific issue. This would come into play during the night shift at the factory where we conducted the user study as a single operator operates a production line, leaving limited options for eliciting help from others.

4.6.3. Limitations

We used the same prompt for all LLMs; however, it is possible that some of the LLMs would perform better with a prompt template developed explicitly for it. Furthermore, we matched the settings of the LLMs (e.g., temperature) as closely as possible across all the tested models; however, the same settings across model types were not completely equivalent, and in the case of Llama 2, we did not have access to the presets as we did not host it locally.

The study's design did not include a real-world evaluation involving end users operating the production line, as this was considered too risky for our industry partner. Such an environment might present unique challenges and considerations not addressed in this research, such as time pressure. However, by involving operators and managers and instructing them to pose several questions based on their actual work experience, we could still evaluate the system and collect valid feedback.

Our benchmarking procedure involved 20 questions, and a singular coder assessed the responses. This introduces the potential for bias, and the limited number of questions may not cover the full spectrum of complexities in real-world scenarios.

4.7. CONCLUSION

This work was driven by the need to overcome the limitations of traditional intent-based conversational AI systems in managing the vast and often unstructured knowledge within manufacturing environments. By leveraging recent advances in large language models (LLMs), we aimed to create a system that efficiently supports operators in their knowledge-intensive tasks, enhancing overall productivity and decision-making.

The results demonstrated GPT-4's superior performance over other models regarding factuality, completeness, and minimal hallucinations. Interestingly, open-source models like StableBeluga2 and Mixtral 8x7B followed close behind. Additionally, the user study highlighted the system's user-friendliness, speed, and logical functionality. However, improvements in the user interface and content specificity were suggested, along with potential new features. Benefits included modernizing factory operations and speeding up certain tasks, though concerns about safety, efficiency, and inferiority to asking human experts were raised. In the following chapter, we explore these concerns and benefits in more depth while highlighting the differing perspectives of operators and their managers.

5

Factory Operator Perceptions of Cognitive Assistants

In Chapter 4, we evaluated an LLM-powered cognitive assistant (CA) in a factory, receiving both positive feedback and areas of concern from factory personnel. In the studies presented in Chapter 2, 3 and 4, we have observed tensions between operators and management regarding the need for CA, how they should be deployed, and their potential impact in the real world. In this chapter, we investigate these topics, paying close attention to the differences between the operator and manager perspectives. Thus, this chapter addresses the gap in the literature regarding the factory operator's perspective on deploying AI-powered knowledge management systems in factories.

Our investigation, which spans over two years and four variations of CAs, examined their usability and effectiveness in real world and lab settings. While we focused on evaluating with factory operators—who are responsible for setting up, operating, and fixing complex production systems—we also involved other stakeholders, including managers and maintenance technicians. Based on the qualitative feedback we collected during the deployments of CAs at two factories, we conducted a thematic analysis to investigate the perceptions, challenges, and overall impact on factory operation work and knowledge sharing. Therefore, this chapter presents a summative and longitudinal study providing comprehensive insights gathered over multiple years.

Our results indicate that while CAs have the potential to significantly improve efficiency through knowledge sharing and quicker resolution of production issues, they also introduce concerns around workplace surveillance, the types of knowledge that can be shared, and shortcomings compared to human-to-human knowledge sharing.

This chapter is based on a manuscript currently under review: S. Kernan Freire, T. He, C. Wang, E. Niforatos, and A. Bozzon. "Operators' Perspectives on Conversational AI for Knowledge Sharing: Challenges, Risks and Impact on Work". Submitted: Proceedings of the ACM on Human-Computer Interaction (CSCW 2024).

5.1. Introduction

Recent advances in Natural Language Processing (NLP), such as Large Language Models (LLM) like GPT-4 [208], have drastically increased the potential of cognitive assistants (CA) to support knowledge-intensive work. A foundational LLM can offer various functions, such as answering general knowledge questions, refining text, and checking logic. Additionally, the information contained in the base models can be extended by providing it with context material, a process called Retrieval Augmented Generation (RAG) [99]. LLMs are capable of more sophisticated, flexible, and human-like reasoning than previously possible with AI. These attributes could help overcome issues with agential AI during work support, such as not considering the social and context aspects of knowledge sharing [80, 101].

Recent technological advances, such as significantly improved natural language understanding (NLU) and AI reasoning demonstrated by LLMs (e.g., GPT-4 [208]), have enabled the development of more intelligent cognitive assistants (CA). However, their application in factories is in its infancy. Working on a production line can be noisy, high-stakes, and dynamic. As such, it is a challenging environment to introduce CAs, as shown by the challenges other digital technologies have faced (e.g., Augmented Reality) [209]. Prior work has explored the barriers and challenges associated with making the step toward smart manufacturing in general (e.g., [210–212]), but not CAs for knowledge sharing on the shop floor, nor thorough real-world deployments integrating the latest technology advancements such as LLMs. This chapter addresses this gap.

Over a period of two years, we deployed CAs in four phases, ranging from technology probes simulated by researchers to voice assistants connected to a live production line to LLM-powered chatbots. Using feedback from system evaluations and responses from semi-structured interviews, we present findings from a hybrid deductive/inductive thematic analysis of operators' and management's perceptions of CAs conducted at two factories. Fortunately, unlike prior work [213], we could evaluate our prototypes and interview many operators without managers present. As such, this work explores the socio-technical and practical factors surrounding deploying CA systems in the factory context from both the operator's and managers' perspectives, leading to the following dissertation research question: **What are factory operators' and management perceptions of the impact, and socio-technical risks and challenges of using CAs for knowledge sharing?** (RQ4).

Through our investigation, we aim to contribute a better understanding of cognitive assistant deployment in manufacturing, providing a foundation for the design and implementation of future conversational AI knowledge sharing tools that are both effective and sensitive to the socio-technical context of operating complex systems in factories. Our contributions can be summarized as follows:

1. We provide a deeper understanding of the opportunities, challenges, and risks associated with using (LLM-powered) cognitive assistants by factory operators from both operator and management perspectives.
2. We provide a deeper understanding of the tensions between factory operators and management regarding technology-facilitated knowledge sharing.

3. We present design guidelines for effective (LLM-powered) cognitive assistants in manufacturing settings.

5.2. Background and Related Work

This section provides an overview of current research in Knowledge Management (KM) in manufacturing and the utilization of technology, specifically cognitive assistants, to enhance knowledge sharing in organizational settings. It discusses the challenges and advancements in KM practices, the significant role of technological interventions in facilitating knowledge acquisition, dissemination, and application, and the socio-technical aspects that influence implementing these systems in manufacturing environments.

5.2.1. Technology for Knowledge Management

Organizational knowledge management (KM) involves leveraging the knowledge of employees, customers, and suppliers [7]. This entails processes to acquire, manage, and share knowledge within an organization to achieve a competitive advantage [8]. In practice, the core tenet of KM is to save information, we might need it later [40]. For this work, cognitive assistants were deployed to support knowledge sharing between factory operators.

Successful technology use for a KMS is demonstrated in Xerox's Eureka project from the 1980s [53]. They developed a remotely accessible system to store and retrieve service call reports from their customer service engineers [53], saving the corporation \$100 million [52]. Xerox's EUREKA development team involved the customer service engineers throughout the development process, inspiring local "champions" to push for change, resulting in a shift in corporate culture. This echoes factors that have been found to lead to successful KM initiatives, such as identifying a distinct group with a particular need for knowledge [214], a well-defined and committed approach to knowledge management [215], along with the backing of top management and the necessary technical and organizational frameworks [216]. As such, KM requires a concerted effort, supported by technology, and is deeply intertwined with social factors.

While Eureka was an early example of an online knowledge base, expert systems were the first widely used AI systems to reason over a knowledge base to support human decision-making. Using AI assistants for KM has progressed significantly from early expert systems in the 1970s, such as Dendral [217] and MYCIN [218], which were designed to answer questions in specific domains based on rules defined by domain experts. In recent years, the improvements in natural language processing and deep learning have given rise to more sophisticated AI assistants, such as IBM's Watson [219] and systems using LLMs (e.g., GPT-4 [208]). These AI-powered systems can handle a wide range of tasks by understanding and generating human-like text.

Foundational LLMs do not contain context-specific knowledge—for example, how to fix a machine in a specific factory—they possess extensive general knowledge and reasoning abilities [88], enabling them to excel in processing complex information [87], generate insights and reasoning [89]. While an LLM can be trained from the ground up for a specific domain, such as medicine, this is typically cost-prohibitive for individual organizations. Even so, using a domain-specific LLM would have two key challenges: (1) reliance on outdated information from their pre-training data, and (2) potential inaccuracies in factual content, a phenomenon termed "hallucination" [88, 95]. To mitigate these issues and maximize the utility of LLMs in specialized, knowledge-intensive tasks, techniques such as fine-tuning, chain-of-thought [96] few-shot prompting [97, 98], and retrieval augmented generation (RAG) [99] can be adopted. RAG involves retrieving relevant information from a knowledge base before generating an appropriate response for a human. RAG is an efficient way to harness the advanced NLP and reasoning abilities of LLMs to answer people's questions on domain-specific topics without needing to train a new model from scratch. Furthermore, the availability of source material—the information retrieved from the knowledge base—improves transparency and enables fact-checking. As such, foundational LLMs used in conjunction with knowledge bases hold significant potential to enhance organizational KM.

5.2.2. AI Assistants for Factory Operators

Recent technological advancements, particularly in natural language processing (NLP), has driven research into AI assistants for factory operators. These systems aim to enhance decision-making, operational efficiency, and training [19, 27, 56–61]. Despite the initial lack of focus on social and human aspects [27], the Industry 5.0 paradigm has brought human well-being to the forefront [220, 221]. Factory AI assistants are part of a socio-technical system that includes the operator, their tasks, and the enabling technology [62]. These assistants aid in problem-solving, machine setup, information retrieval, remote control, decision support, and knowledge sharing, often through conversational AI interfaces [60, 63–65]. Integration with factory systems like control and monitoring systems, scheduling systems, and sensors enhance their capabilities, with some employing machine learning for predictive maintenance [66] and NLP for knowledge extraction [67]. However, the knowledge bases informing these assistants are typically static, defined by domain experts [19, 21, 68, 69], with limited live data usage focused on presenting production status and performing basic reasoning [20, 70–73]. Examples include AI assistants for operator training [19], smartwatches for error notifications [73], and VR-enabled consultation systems [21]. While these systems effectively deliver knowledge and present information, they often overlook acquiring knowledge from operators.

5.2.3. Cognitive Assistants: Acquiring and Sharing Human Knowledge

In the context of this dissertation, **cognitive assistants are an advanced type of AI assistant that interacts conversationally to acquire knowledge from humans, share it with other humans, and support its application.** Cognition refers to the mental processes of acquiring and comprehending knowledge [74], which the CA aims to support. To do so effectively, a CA relies on several technologies, including dialogue management, NLP, context awareness, databases, and ontologies. While deploying

CAs at a production line appears novel, we discuss notable works that evaluated (parts of) the underlying technologies and techniques in the following sections.

Acquiring knowledge using conversational AI and/or mobile devices has been explored and shown to be promising. For example, Fenoglio *et al.* [75] introduced a role-playing game involving virtual agents, human experts, and knowledge engineers to refine knowledge graphs that were algorithmically generated. The authors raised several important ethical concerns regarding the processing of personally identifying data, electing to avoid audio and video recordings, and other concerns regarding misuse of employee tracking. Other examples outside of manufacturing demonstrated the successful use of conversational AI to crowdsource knowledge acquisition, namely, a game to elicit knowledge from crowdoperators Balayn *et al.* [76] and a context-aware chatbot that simultaneously fulfills information needs while acquiring new knowledge [77]. A key consideration for these systems is ensuring the quality of acquired knowledge, thus employing rating or validation mechanisms. Although these conversational AI systems were not been deployed in a live production environment, they demonstrate the potential of using conversational interactions as an efficient means to acquire knowledge.

The system presented by Hannola *et al.* [27] is closely related to our systems, except for the lack of conversational AI. It comprises a mobile application to take notes and videos of solutions to production line problems, enabling others to learn from newly created knowledge [27]. While most operators liked the system, there were exceptions, citing feelings of being controlled, lack of available knowledge, and difficulties typing reports. The authors concluded that documenting solutions immediately on the shop floor was ideal, noting the importance of minimizing the documentation workload. Additionally, the original knowledge contributor is openly credited so that operators can consider this in their decision making process or initiate a conversation regarding the solution. Overall, the operators believed the system helped increase performance and the quality of their work, indicating the potential of supporting knowledge sharing through mobile devices on the shop floor.

5.2.4. Socio-Technical Challenges

While many KMS have been deployed in manufacturing, the human potential for knowledge sharing and the associated social factors have been largely overlooked [27]. Work in factories is affected by social and organization factors rather than purely technical, deterministic activities [78, 222]. While it is clear that digital tools can facilitate knowledge sharing among users, for example, by enabling factory operators to post comments under existing work instructions to facilitate discussion about best practices [79], these tools can have many unforeseen social effects and barriers. For example, introducing relatively simple tools, such as a machine repair ticketing system, can shift the balance of control towards or away from operators depending on who is given privileges to create, view, and modify tickets [78]. Furthermore, operators omitted their valuable informal knowledge when asked to create learning material collaboratively in favor of the standard working procedures [223]. Additionally, requiring operators to use smartphones to document their work can be perceived as controlling [27]. Thus, these factors emphasize the importance of considering social factors when introducing new digital knowledge sharing practices, which are generally overlooked when designing AI assistant systems [80].

To tackle the complex social challenges discussed above, it is crucial to look beyond the technical aspects when developing AI tools for knowledge sharing. Although not social in nature, considering human-computer interaction factors such as avoiding information overload and ensuring the information needs are met on an individual level are also key [5]. Indeed, recent work on assistance systems for sharing tacit knowledge in factories that had focused on technological aspects emphasized the importance of comprehensively evaluating the social, psychological, and organizational implications [58]. The crucial importance of socio-technical factors is what motivated us to holistically reflect on the feedback we received from operators and managers while deploying cognitive assistants, culminating in this longitudinal study.

5.3. Real-World Case Studies in Factories

5.3.1. Approach and Context

We conducted a hybrid deductive/inductive thematic analysis of 251 comments collected from operators and factory management over two years of CA evaluations at two factories in two European countries. The factories manufactured detergents on production lines that could rapidly switch between over a hundred detergent types depending on customer demand. The factories operated 24/7 with three eight-hour shifts per day, and each production line was typically manned by two operators. The operators reconfigure, operate, and resolve issues with the production lines, all knowledge-intensive tasks, especially considering the frequent production and quality issues. We evaluated four CAs of varying levels of sophistication and functionality, from technology probes simulated by researchers to CA systems connected to a live production line, allowing us to explore many facets of the operators's experiences. This research was approved by the factory workers' councils and our institution's human research ethics committee.

5.3.2. Systems, Protocols, and Participants

The development of the CAs used in this study was initiated according to objectives set by the factory management: to increase production performance and reduce training time by supporting (tacit) knowledge sharing amongst operators and training operators to follow standardized procedures. The factory management had observed significant disparities between operator shift performance, poor adherence to standard working procedures, and inefficient human-human knowledge sharing practices. Despite the top-down initiation of the project, we involved the operators throughout the development and evaluation of the system to ensure that we were also meeting their needs and values. The evaluated functions are listed in Table 5.1 and details of the *N=40* participants in Table 5.2. In the following sections, we describe four phases of data collection, including the capabilities of the CAs, data collection protocols, and participant information.

No.	Capability	Ph. 1	Ph. 2	Ph. 3	Ph. 4
1	Speech input	✓		✓	✓
2	Speech output	✓		✓	
3	Text input		✓	✓	✓
4	Text output		✓	✓	✓
5	Capture issue handling knowledge	✓	✓	✓	✓
6	Share issue handling knowledge		✓	✓	✓
7	Automatic validation of captured knowledge				✓
8	Human approval of captured knowledge				✓
9	Capture product knowledge		✓	✓	
10	Rate captured product knowledge		✓	✓	
11	Approve shared product knowledge		✓	✓	
12	Share product knowledge		✓	✓	
13	Graph machine data for systematic reflection			✓	
14	Machine vision for context-aware dialogue		✓	✓	
15	Upload new documents to the knowledge base				✓
16	Answer predefined FAQs		✓	✓	
17	Provide relevant standard work instructions		✓	✓	
18	Answer queries using factory documentation				✓
19	Real-time explainable machine settings advice			✓	

Table 5.1: CA capabilities evaluated per phase.

Phase One: Technology Probe during Production Line Operations

In preparation for the first evaluation phase (Jan. 2021 – May 2021), we asked the operators (*n*=2) and their managers (*n*=2) to describe the current knowledge management practices at the factory. Together, we defined the information needed to represent an operator's knowledge regarding issue handling, namely, the symptoms, machine component(s) associated with the symptom, type of issue, location, current task, solution, and root cause analysis. During the first phase of evaluations, we used a technology probe [224] to evaluate the user experience of using a CA for capturing knowledge at the production line.

Over three days in May 2021, we asked factory operators to share how they solved issues with a CA being simulated by a researcher. Participants used a Bluetooth headset to interact verbally whilst at the production line. The researcher simulated the CA using a protocol representing state-of-the-art NLU and dialogue management. The protocol followed a form-filling conversation during which the operator was asked to describe the issue according to the abovementioned representation. If the operator missed specific information or used pronouns (e.g., *it*) to describe components, the researcher would ask for clarification. Data was collected over three days, and six (*N*=6) male operators participated. Additionally, the user study was partially observed by two managers and a technician. Throughout the user study,

Table 5.2: Participant Data

No.	Role	Factory	Phase	Test duration (min)	Prior Phases	Gender
1	Operator	1	1	60-180	No	Male
2	Manager	1	1	0	No	Male
3	Operator	1	1	60-180	No	Male
4	Manager	1	1	0	No	Male
5	Operator	1	1	60-180	No	Male
6	Operator	1	1	60-180	No	Male
7	Technician	1	1	0	No	Male
8*	Mixed	1	2	10	No	Mixed
21	Manager	1	3	15-30	Phase 2	Male
22	Operator	1	3	15-30	No	Male
23	Technician	1	3	15-30	No	Male
24	Operator	1	3	15-30	No	Male
25	Operator	1	3	15-30	No	Male
26	Manager	1	3	15-30	Phase 2	Male
27	Manager	2	4	15-30	No	Female
28	Manager	2	4	15-30	No	Female
29	Manager	2	4	15-30	No	Male
30	Manager	2	4	15-30	No	Female
31	Operator	1	4	15-30	No	Male
32	Operator	1	4	15-30	No	Male
33	Manager	1	4	15-30	Phase 2,3	Male
34	Operator	1	4	15-30	No	Male
35	Manager	1	4	15-30	Phase 1,2,3	Male
36	Operator	2	4	15-30	No	Male
37	Operator	2	4	15-30	No	Male
38	Operator	2	4	15-30	No	Male
39	Operator	2	4	15-30	Phase 3	Male
40	Operator	2	4	15-30	Phase 1	Male

researchers recorded observations by the managers and operator feedback regarding the operators' experiences and challenges during interactions with the CA, resulting in 18 (n_c=*18*) comments relevant to using CAs in factories. Key findings from the user study include the negative impact of factory noise on voice interaction, discomfort from wearing headsets for prolonged periods, and the need for shorter interactions when eliciting information from operators.

Phase Two: CA with a Simulated Production Line

Having learned the importance of efficient and reliable interactions from the first phase of evaluations, we developed a CA that used live data from a simulated production line and human location tracking using stereoscopic cameras[1] for context

[1] https://www.stereolabs.com/products/zed-2–last accessed November 20, 2024.

Figure 5.1: Headset used at the production line for data collection during phase 1

awareness. A computer vision model converted the camera stream into 18 XYZ points representing an anonymous operator skeleton. We built a section of a production line in the lab, including human-machine interfaces and a computational model that simulated the machine's behavior to develop the location tracking capabilities in a stable environment (see Figure 5.2). Then, we installed the stereoscopic cameras at the factory. Context awareness enabled the text-based CA to facilitate efficient user interactions by prefilling user responses as buttons depending on their proximity to machines. For example, during knowledge sharing about issue handling or machine setup, the CA would use the live data to suggest error codes or machine components and ask the operator to confirm. In addition to enabling conversational shortcuts, we reduced the required information for an issue report, giving the operators more freedom to choose how detailed the report needs to be.

As voice interaction was unreliable due to factory noise and frequent use of factory-specific jargon, we opted to develop a text-only CA for phase two (see Figure 5.3). The CA was built using version 1.0.6 of RasaX[2], an open-source framework for building context-aware conversational agents. The assistant used an action server to connect to additional services, most importantly a Neo4j[3] knowledge graph for storing the captured knowledge.

We conducted three focus group sessions, totaling 11 (*n=11*) factory operators and managers, two (*n=2*) of which identified as female, and nine (*n=9*) as male. The sessions consisted of the following four activities: watching a video presenting the simulated production line, a video presenting the CA and how to use it, a structured user test that involved completing eight tasks with the CA, and finally, a semi-structured user test where participants were free to experiment with the CA,

[2] https://legacy-docs-rasa-x.rasa.com/docs/rasa-x/1.0.x/–last accessed November 20, 2024.
[3] https://neo4j.com/–last accessed November 20, 2024.

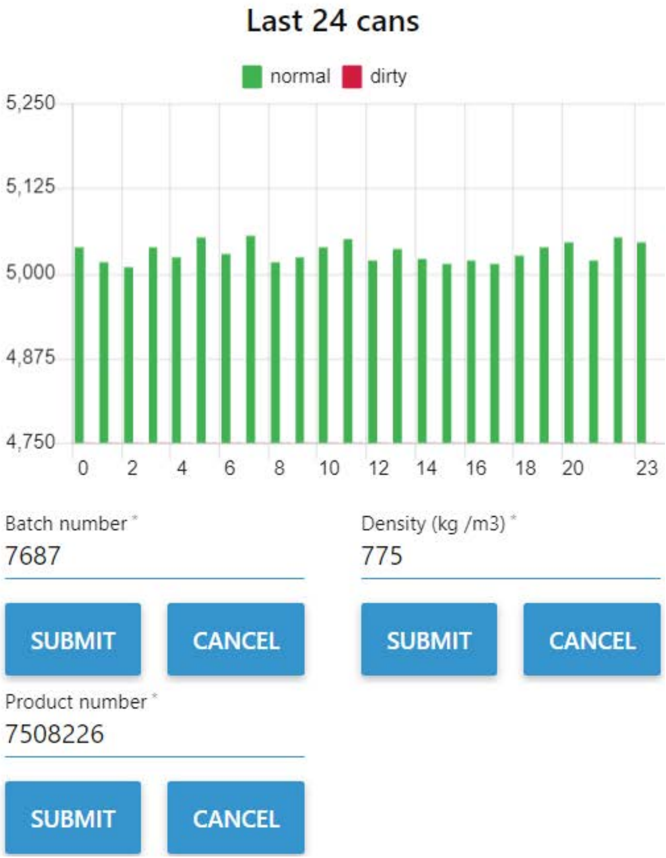

Figure 5.2: A simulated "weight checker" user interface that was connected to the assistant during phase 2. Each bar represents the weight of a canister that was filled with detergent. The "weight checker" weighs each canister and stops the production line if the weight falls below a specified threshold.

evaluating its capabilities in supporting operators. The tasks for the user test focused on retrieving and providing information to the CA aided by live data from the simulation but did not require the participants to act upon the retrieved information; for example, we instructed the participants to "Find instructions on how to perform a prerun", "Record the current filling machine settings" and "Ask for help with this problem: 'Symptom: The filler is foaming. Error code: 33. Product: 7508226."') At the end of each activity (e.g., after completing all eight tasks), participants were asked to write post-it notes on their experience and how it could be improved. Additionally, we facilitated group discussions for which a second researcher took notes. The anonymously collected (post-it) notes resulted in 55 (n_c=55) comments that were relevant to this study. We excluded comments that were not directly related to the CA, for example, if they were solely related to the realism of the production line

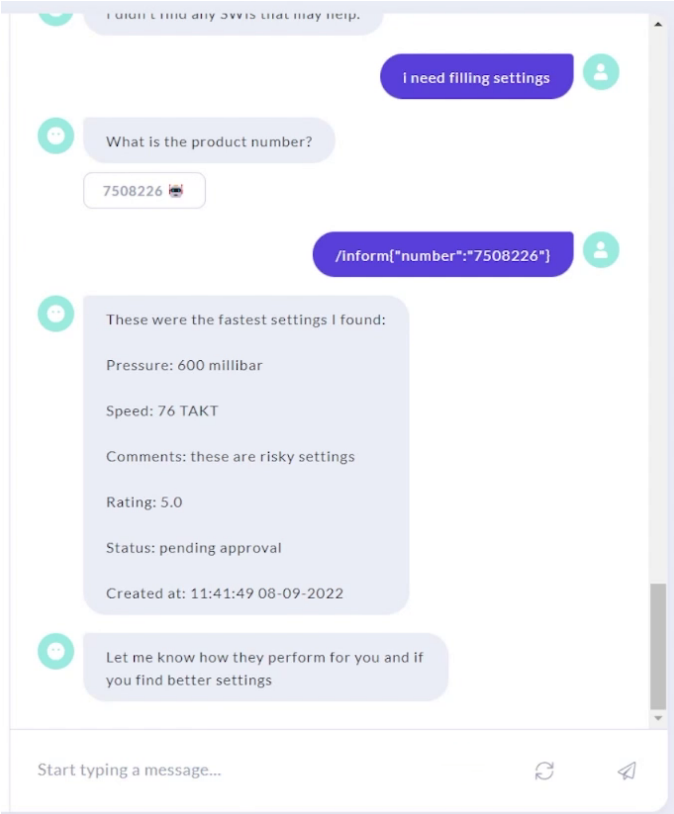

Figure 5.3: The context-aware chat interface used during phase 2.

simulation. The participants from phase two are referenced collectively as P8* as we did not track comments to individuals.

Phase Three: CA with Partial LLM Integration Connected to the Production Line

During this phase, we evaluated a CA capable of voice and text interaction with access to live machine data from the production line. In addition to context-aware user interactions, the data was used to generate statistics and graphical representations of the production line performance and operator operations. The objective was to use these visualizations to support work practice reflections and knowledge discovery. Additionally, the CA used an LLM-powered RAG function to answer user queries using the GPT-3.5 API (version gpt-3.5-turbo-0301[4]) and a corpus of shared operator knowledge. The CA, which used the Rasa framework[5], was accessible via an Android

[4] https://platform.openai.com/docs/models/gpt-3-5–last accessed November 20, 2024.
[5] https://rasa.com/docs/rasa/–last accessed November 20, 2024.

application on a smartphone based on the Mycroft companion app[6](see Figure 5.4.

Figure 5.4: Smartphone user interface for phase 3 and 4

At this stage, we demonstrated the capabilities of the CA to five (*n=5*) operators and three (*n=3*) managers. All participants identified as male. The operators were given the opportunity to use the CA at the production line as they saw fit (e.g., to record and retrieve information about recent issues), whereas the managers were asked to complete five tasks in an office setting. We asked the managers to complete the following tasks as if they were operators: save and retrieve machine parameters and associated tips, save and retrieve a solution to a problem, generate production statistics for the past 12 hours, and finally, pose a question about a problem using the LLM-powered response generator. Following this, we conducted semi-structured

[6] https://github.com/MycroftAI/Mycroft-Android–last accessed November 20, 2024.

interviews regarding their perceptions of its knowledge sharing capabilities, user interactions, benefits, risks, and suggestions for improvements. We extracted relevant statements from the interviews, resulting in 60 (n_c=60) comments for this study.

Phase Four: LLM-powered Conversational AI

During the final evaluations, we tested the LLM-powered functions exclusively, using the gpt-3.5-turbo-0613 model API[7], LlamaIndex[8] for the response generation and Gradio[9] for the frontend (see Figure 5.5).

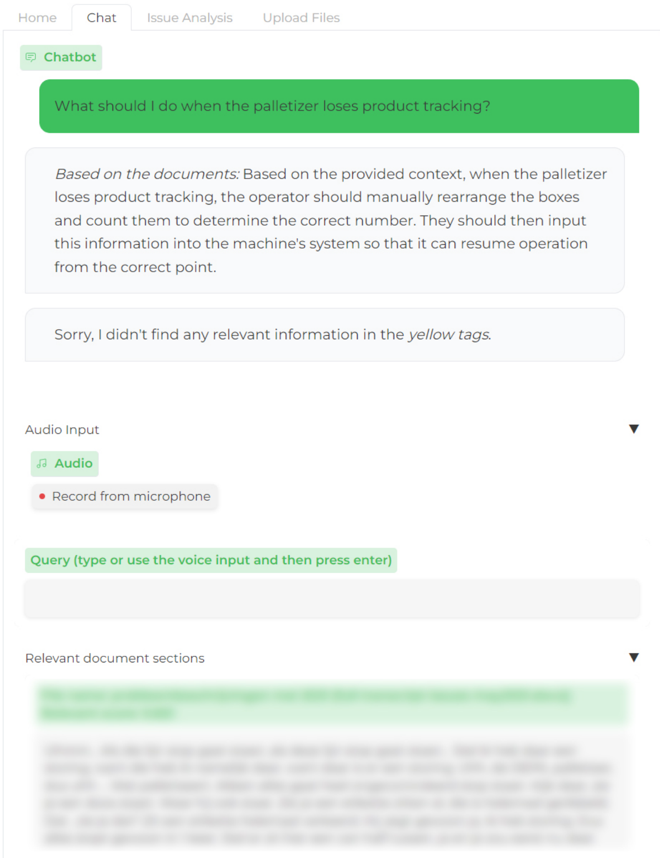

Figure 5.5: The chat tab of the LLM-powered CA from phase 4

The LLM-powered system introduced the following new functions: RAG for user queries based on a database of operating documentation and submitted root-cause analysis reports; a digital form for submitting root-cause analysis reports; automatic

[7] https://platform.openai.com/docs/models/gpt-3-5–last accessed November 20, 2024.
[8] https://docs.llamaindex.ai/en/stable/–last accessed November 20, 2024.
[9] https://www.gradio.app/–last accessed November 20, 2024.

validation of submitted root-cause analysis reports; human validation of submitted root-cause analysis reports; and a function to upload new operating documentation. We evaluated the LLM-powered CA with three (*n=3*) operators and one (*n=1*) manager from the first factory, and four (*n=4*) managers and five (*n=5*) operators from the second factory. The evaluation in the first factory was conducted in person at a production line, with one researcher coordinating the evaluation and conducting the interview and a second researcher recording audio and taking notes. Conversely, the evaluation in the second factory was conducted asynchronously without the presence of a researcher. The operators and managers from the second factory were provided with an online survey that specified the tasks and open-ended questions that were used to extract the comments. The tasks for both factories aimed to evaluate the knowledge capture and sharing capabilities of the CAs with recent issues from their work. For example, we instructed the participants to ask for help with recent problems they faced on the production line. Similarly, we asked them to record a solution and their thought process for a problem they recently faced. After completing the tasks, the open-ended questions for both factories included the following topics: the perceived benefits, risks, adoption barriers, operator-management relations, and suggestions for improvements, resulting in 118 (n_c*=118*) comments.

5.3.3. Hybrid Deductive/Inductive Thematic Analysis

After several rounds of qualitative data collection over two years, we conducted a hybrid deductive/inductive thematic analysis with two independent coders as described by Fereday and Muir-Cochrane [106] and recently applied by Martinez-Maldonado *et al.* [225]. The analysis was conducted based on 251 (n_c*=118*) comments we extracted from interviews, survey answers, and post-it notes that were collected during phases one through four. All comments related to the user experience, social, technical and organizational factors of the CA deployments at the factories were included. Conversely, unintelligible comments were excluded. Complex statements comprising multiple standalone arguments or observations were divided into separate comments to facilitate sorting and analysis. If the other part(s) of the original statement provided relevant context for the divided comment, this was included in regular brackets.

For the thematic analysis, the first level involved a deductive approach to define themes closely related to the research questions and literature review, resulting in the following six themes: (1) Impact on Work Experience, (2) Optimizing Knowledge Sharing, (3) Adoption and Change Management, (4) Privacy, Safety, and Ethics, (5) Usability and User Experience and (6) Technical and Operational issues. The second level followed an inductive approach to identify subthemes by two independent coders. Finally, the subthemes were collaboratively revised to reach a consensus, resulting in 20 subthemes. At this stage, we resolved all intercoder discrepancies. At both levels, we allowed comments to be sorted into multiple themes and subthemes to accommodate a holistic understanding of participants' perspectives on real-world issues.

5.4. Operator and Management Perceptions

5.4.1. Impact on Work Experience

Improved Information Retrieval

The comments suggest that the system positively impacts information retrieval within the organization. The CA was perceived to provide immediate answers to doubts about machine operation, making access to company documents more user-friendly (P30), and improving the flow of information (P28). Furthermore, participants saw it as a tool that modernizes the factory (P27). However, one concern is that operators may benefit more from traditional information retrieval systems (P40).

Useful for Problem-Solving

Many participants (*n*=9) are positive about the usefulness of the system and specific functions (P22, P24, P25, P36, P30, P31, P32, P33, and P34), and some participants, such as P25, make a distinction between some functions being useful and others not. Specifically, the system provides greater speed in carrying out some small tasks (P29), it has a well-set-up logical control function (P30), and helps solving problems quickly (P8*, P24, P25, and P37) as stated here "I find it nice that you can look back. What did we do last time with the same problem? That, I think, is a significant advantage." by P24, and "[the CA] can be a great tool in problem solving in the day-by-day activities." by P8*. However, there is a disadvantage of wasting time looking for a solution to a problem if it is not reported in the system's history (P29).

Effective Knowledge Sharing

Several benefits regarding knowledge sharing are mentioned by the participants, namely that it is useful to transfer knowledge from experienced to novice operators (P31), "The system also allows for the communication of tacit knowledge between teams." (P33), it can also help individuals recall knowledge of how they previously solved issues (P24, P32) and experts' knowledge can now persist and is always available (P34). These comments show that participants generally see it as a mechanism for experts to share their knowledge with novices and as a memory augmentation for themselves.

That said, there are concerns that receiving information from the system could be less quick, effective, and comprehensive than asking the expert staff present (P27). This is likely true, assuming that there is an expert present with the relevant knowledge. However, our observations at the production line and statements by operators, such as "A few hours are lost every week on previously solved issues." (P33) and P34 stated that operators "Lose about half an hour per day over one team per line." suggests that information on previously solved issues was not readily available.

Whereas management believes in standardizing approaches so every operator operates at the highest achievable performance, some operators do not think this is feasible. One operator in particular, P25, was skeptical of the benefits of knowledge sharing as stated here, "We always adjust to the specific product characteristics and line conditions.", "We often have different issues." and "Because everyone has their own approach, some run slower, and others have their own techniques.”. In other words, their work is so complex that the ‘best practice’ is highly dynamic and dependent on context factors and individual strategies. This reflects experienced operators’ pride in the strategies they developed, which may hinder the effectiveness of knowledge sharing. Conversely, P22 mentioned that “It would be useful to know if and how other operators have produced at higher speeds.” (P22), suggesting they would be interested in learning from each other. This reflects a split among operators regarding the perceived usefulness of knowledge sharing.

Training Time

Opinions are divided on whether training time will be reduced, with P32 stating that “[the CA] will address staff shortages and training time.” whereas P31 stated that training will not directly be reduced but independence will be increased. Others, such as P33, were unsure whether training time would be impacted but expect improvements as the technology matures. Generally, participants thought that novice operators would benefit most (P8*, P22, P24, P25, P31, P34, and P38); for example, P22 stated, “This is very nice, yes, especially for beginner [operators].” and P25 stated, “It’s useful for newcomers who might not have as much product and line knowledge.” These statements support the idea that independence can be improved and training time can be reduced.

5.4.2. Optimizing Knowledge Sharing

Value of Efficient Knowledge Sharing

The comments highlight the importance of operator knowledge at work, confirming the importance of capturing this knowledge. This know-how is not captured in manuals or documentation; for example, “An operator poured water over one of the canisters in the filling machine when the line had stopped to get it moving again.” (P6) and “We adapt to the product’s characteristics, and some products require slower operation due to foam production.” (P25). Yet, existing mechanisms for capturing this knowledge were largely unused, as stated by P2: “Because the problems are so poorly documented in [current issue reporting system], we don’t have a good overview of the problems.” P25 also describes how they develop their own strategies: “Because everyone has their own approach, some run slower, and others have their own techniques.” and the necessity to adapt to changing context: “Yes, sometimes we have products that change their speed, or the machine is modified, allowing faster operation.” This partially explains why existing tools for documenting knowledge are quickly outdated and resource-intensive to maintain, as discussed in section 5.4.2 below. Overall, these points demonstrate the high potential value of an efficient knowledge sharing mechanism.

Knowledge Representation

There were some criticisms of the generic nature of (some of) the information provided (P39). Related to this, there were suggestions for improvement, such as more detailed, precise, and updated information (P27, P37, P40) and more specific instructions (P40). This suggests that knowledge representation should be comprehensive and precise, focusing on providing detailed instructions or information where necessary. Indeed, participants were concerned that some knowledge would be too complex for the system; for example, “Problems at the production line can be more complex and have complex solutions.”; “Mechanical problems can be hard to describe. pictures might be better.”; and “It will be challenging to match similar problems using text only as people can describe them in many different ways.” (P8*). Technically, this was a significant challenge when using keyword-based search and intent-based assistants; however, semantic search and LLM-powered RAG can handle divergent phrasing much better. Other operators believed a simple approach would be sufficient; for example: “Error code and description would be enough information to save.” (P24), and P22 reiterates that “Error codes are the single most important factor to match to existing issue solutions.”. However, P25 points out that “Sometimes there are random issues that don’t have specific fault codes.”, demonstrating a need for alternatives. These wishes reflect a desire to minimize interaction time with the system. Therefore, a tension exists between the perceived effort required to document knowledge and the value of that knowledge for sharing purposes.

Knowledge Maintenance

The comments discuss the necessity and challenges of knowledge maintenance from several perspectives. P40 states that “If the system is not kept updated, it becomes useless.” and P24, whilst reflecting on previous tools for KM, mentioned that “The biggest challenge is keeping it updated. It takes so many man-hours, and you just don’t have those.”. Additionally, participants raised concerns about not having sufficient data, such as P8*, who stated that “[The system] might not give good recommendations if it’s only half full of knowledge.” and “[The system] needs more data.”. There are also concerns about how to share knowledge clearly if there are many (potentially conflicting) entries in the knowledge base to share (P8*).

Some resistance to participating in the upkeep of knowledge is evident as operators reported that uploading documents is not their responsibility (P40). There were also concerns about how the database will be connected and updated with instructions and documents (P8*), and the need for a proper process for approving knowledge updates to maintain high quality, as stated here by P8*: “We will need to have a proper process for the settings change confirmed by an engineer.”. Comments on the existing KM situation at the factory portray the challenges in keeping the knowledge updated, as discussed in section 5.4.3.

5.4.3. Adoption and Change Management

Operator Support and Assistance

The comments suggest that operator support and assistance are crucial for adopting the CA. P29 highlights the importance of awareness and training for operators, which could potentially hinder adoption if not adequately provided. P37 also emphasizes the need for operators to know how to use the system correctly to improve adoption.

OPERATOR ACCEPTANCE AND ADOPTION CHALLENGES

The comments reveal several challenges to operator acceptance and adoption of the CA. Managers are concerned that "If two or three bad experiences occur, then [operators] might be discouraged from using it.". Some participants refer to the need for full adoption; for example, P30 suggests that adoption could be hindered if all employees don't accept the system, indicating a need for broad buy-in. Furthermore, P38 implies that older operators' usage could be a potential barrier to adoption, and P34 also points out that "[User acceptance] Depends on digital literacy of operators.", suggesting that age and technological proficiency may play a role in adoption. However, P34 clarifies that they have no concerns about adoption. P35 reveals privacy concerns as a potential obstacle to adoption, with one operator refusing to participate due to discomfort with human tracking, as discussed in section 5.4.4. This remains a critical challenge for adoption, especially if rogue operators sabotage the system. Other potential barriers to adoption include poor performance (5.4.5) due to lack of details and out-of-date knowledge (5.4.2), poor accessibility (5.4.5), lack of cultural importance for knowledge sharing (5.4.3), and a high perceived effort to record knowledge (5.4.2).

A few participants referred to failed attempts to introduce new documentation tools; for example, P23 points stated: "We bought a bunch of tablets 7 years ago with the intention of using them for documentation, but they were never adopted." and that current documentation tools are poorly used (see section 5.4.3 for more details.) Although we do not know the reasons behind the failed adoption of tablets for documentation, it does reflect how challenging introducing a new KM system can be, and why both management and operators need to commit.

MANAGEMENT AND OPERATOR PERSPECTIVES

Using a CA to share knowledge among operators is largely motivated by the factory management's observations that this was not naturally occurring effectively despite their attempts to stimulate it. This is reflected in the following statement by P3, an operator: "Management has accused me of not sharing my knowledge with others. However, it was due to a miscommunication about what the problem was about.". The comments generally indicate a range of perspectives between management and operators regarding knowledge sharing practices. Some operators—especially those who were very distrustful of management—strongly believe that knowledge is already sufficiently shared and do not appreciate the suggestion that they are hiding it. Regarding documentation, some operators stated that operating the production line is too intensive, leaving insufficient time for reporting. From the management's perspective, they observe operators who are frequently idle and not proactive in sharing their expertise. Regardless of the truth, we observed that the careful introduction of a knowledge sharing tool such as a CA is key to reconciling the differing perspectives. For example, by framing the CA as a support tool for operators to avoid the impression that experienced operators are being forced into sharing their knowledge.

Indeed, many of the managers believed that the CA would need to be carefully introduced to avoid rejection from the operators. P29 suggests the need to sensitize operators and explain the system's advantages, implying a potential gap in understanding or perception between management and operators. P38 and P39, both operators, suggest that perspectives may vary depending on the individual, with some being indifferent to the differences between management and operators. P1 reveals a desire for operators to be heard by upper management, whereas others are satisfied with the influence operators already have and feel insulated from the desires of (upper) management through, for example, the workers' council (P32). That being said, several operators mentioned that the management implements new tools without fully committing. This has led to distrust toward management regarding whether they would commit to the CA. Conversely, the management believes that the tools they introduced work sufficiently well, and the operator's laziness is the problem. Overall, our experience is that most operators were open to new knowledge sharing tools introduced by management as long as they were involved in the process, the benefit was clear, and the management invested sufficiently in delivering a reliable and useful tool.

Regarding the knowledge sharing interactions with the CA, we observed differences in requirements between the operators and management. Management was mostly concerned with collecting sufficiently detailed knowledge and ensuring operators used the tool to encourage standardized best practices. Conversely, the operators were concerned with ensuring the interactions were as short as possible, minimizing the required details and only using the tool when absolutely necessary. While (upper) management tended to focus on an idealistic view of capturing every detail of an operator's knowledge about a problem, the operators tended to seek out the minimally required detail for describing and retrieving problem solutions. Concurrently, the vision of (upper) management was that all operators follow a single 'best practice', whereas some operators preferred to follow their own intuition and personal strategies, at least when it came to tuning the parameters of the production line. When it came to solving critical problems that stopped production, operators were more open to learning from others. Overall, we recognized a stark difference in perspectives between management and operators, both with valid concerns. These factors demonstrate the complex socio-technical challenges of introducing a tool for sharing knowledge.

Integration with Existing Systems and Practices

The comments suggest that integrating the CA with existing systems and practices is a key consideration. P1 revealed that some operators use WhatsApp for internal communication, suggesting that operators were not satisfied with official communication channels. That being said, the operators still used the company-issued phones to request help from technical services or expert operators, many of whom were more senior. The official policy for resolving problems at the production line consisted of several steps of escalation, beginning with the present operators trying to solve the problem themselves. If unable to solve the problem, operators can request help from other (expert) line operators, shift leads or technical services. If a CA were introduced, it could be positioned as the first escalation step, allowing (novice) operators to elicit support before needing to call on their fellow operators, who are likely busy with other tasks. Most operators supported the idea of sharing their knowledge to support each other in this way, especially less experienced operators and those involved with training operators. However, it remains to be seen if the operators would be willing to maintain the effort for a prolonged period of time.

At the production line, we noticed a pattern where operators frequently skipped or improperly executed tasks they deemed non-essential. Operators revealed that some new procedures are not always followed, such as autonomous maintenance procedures (small maintenance tasks that operators are expected to routinely conduct): "[The autonomous maintenance] cards are frequently turned around without the task being completed." (P3). Furthermore, there is evidence that operators do not properly input information into existing documentation systems, as stated by P2: "Because the problems are so poorly documented in [issue reporting system], we don't have a good overview of the problems." (P2) and "Existing digital report system is frequently unused." (P5). When we asked the operators why they were not using the issue reporting system, we frequently heard that it was unreliable, suggesting that the operators did not see the point of putting in effort to use something that could not be relied upon. Furthermore, the upkeep of standard working procedures has not been performed as mentioned by P24 here: "No, no, but in fact, the current standard working instructions are from 2003 or 4 when we started with it. It has become such a tangled mess. We really want a different system for it. It's all still in Excel." All of this points to a lack of investment in knowledge management systems from the management and a lack of engagement in responsibilities that are perceived as non-essential to operators, possibly reflecting a work culture where operators are disinclined to conduct tasks beyond the operation of the production line. Interestingly, P33 mentioned that "Previous non-digital version existed and was popular but more cumbersome.", indicating that engagement with a previous knowledge sharing system used to be higher but was hindered by its unwieldiness. While the introduction of a CA alone will likely not change the underlying culture directly, it could reduce the perceived effort in documentation and improve the perceived benefit of documentation.

5.4.4. Privacy, Safety and Ethics

Concerns over Operator Privacy and Data Handling

The comments reveal mixed opinions regarding user privacy and data handling. Some users, as indicated in comments P31, P32, P33, and P34, do not express any privacy concerns, with P33 stating, "No [privacy concerns as] most of the data isn't personally identifiable." However, P35 reveals a contrasting perspective, where an operator refused to participate due to privacy concerns and was uncomfortable with human tracking and covert management oversight. P34 suggests a balanced approach, advising the provision of individual profiles for operators but also cautioning about privacy. This suggests that while some users are indifferent or unconcerned about privacy, others are acutely aware and concerned about how their data is handled and monitored.

Security Measures Against Misinformation or Malicious Use

There are some concerns regarding the security risks if the system's advice is incorrect, as stated by P27: "Risks: if the responses are not adequate, you risk safety.". Safety becomes especially critical in inherently dangerous industries such as detergent factories, the context of this study. This raises questions regarding who is accountable for mistakes, especially if the system relies on knowledge shared by other humans, and what level of accuracy is acceptable for the system if human safety is at stake. Furthermore, P33 mentioned a "Small possibility exists that users may be malicious and enter falsehoods but unlikely.". The above points support the necessity of mechanisms to safeguard the quality of the knowledge, such as knowledge input approval procedures, checking for conflicts with existing knowledge, and identifying outdated knowledge, many of which could also be supported by AI (see also section 5.4.6). Furthermore, framing the system as a supporting tool, not a decision-maker, could help remind operators that they are still accountable, as suggested by P34: "Will be useful as a supportive tool but not a solution".

Operator Autonomy and Empowerment

Participants emphasize the importance of operator autonomy and see the CA as a support tool, as stated by P33: "[It will benefit operators] as long as it remains a support tool and operators remain the main actor in the work.". P34 suggests giving each operator their own profile to aid knowledge management, indicating a desire for operator autonomy and empowerment. However, the same comment also cautions about privacy. Indeed, as discussed above in section 5.4.4, P35 mentioned that an operator refused to participate due to privacy concerns about the camera system, underscoring the importance of operator autonomy. On this topic, operator opinions are divided, as P5 states, "No issues with having cameras and would rather that [management] invest in enough cameras to cover the entire line instead of half-measures.", indicating a desire for management to fully commit to a solution that can support operators. Throughout the process, we respected operator autonomy by involving the workers' council in the design process. They approved the work we conducted, and everyone was given the opportunity to ask questions and discuss concerns. However, despite formally approving the deployment of a camera system, at least one operator sabotaged the installment, indicating disagreement indirectly, as mentioned by P21: "[Operators] have placed a sticker over the camera, blocking its view." (see Figure 5.6 below). This highlights an ethical dilemma that organizations and researchers may face when introducing AI systems that tracks humans: operators may not feel safe to oppose the introduction of new tools openly.

5.4.5. Usability and User Experience

User Interface Design and Accessibility

Operators highlighted the importance of a user-friendly, accessible system that caters to diverse user needs. Operators expressed a desire for a more intuitive information display, with one suggesting to "Display the attachments after the main text for easier reference." (P28). Similarly, several suggestions highlight the shortcomings in

Figure 5.6: A sticker used to block the (stereoscopic) camera overlooking the production line.

the systems NLU, for example, requests for recognizing keywords to provide relevant suggestions and streamline communication (P8*). The use of visual aids was also highlighted, with suggestions to “use more pictograms and pictures to speed up understanding.” (P8*). Language localization was another aspect that users found important, with requests for system versions in other (local) languages. Mobile accessibility was appreciated, but concerns were raised about the use of personal devices, with one user questioning, “Are we not going to force people to use their phone for that?”(P24) These comments highlight the need for a system that is easy to navigate, visually intuitive, available in local languages, and providing company devices.

System Performance and Intelligence

The topic of reliable and efficient user interaction is frequently mentioned. Whilst nine (*n=9*) participants are positive regarding the system’s natural language understanding, eight (*n=8*) participants expressed concerns or areas for improvement, highlighting the challenges for user satisfaction. Suggestions include the ability to understand user queries based on keywords (P8*), including autocomplete in the chatbot to speed up the performance, and better NLU when asking for assistance during machine troubleshooting (P8*). There were also calls to load more (detailed) data and knowledge into the system (e.g., P27) to improve performance. Despite these concerns, *n=9* participants provided positive feedback on the system’s performance and usability, suggesting that it is generally well-received and effective in its current form. Two participants appreciated the response speed; for example, P29 stated: “Quick in responses.”. Others highlighted the desire for “More consistent performance.” (P33). The emphasis on reliable, speedy performance is logical, considering one of the main perceived advantages of the system is to speed up problem solving as stated by P37: “Advantage: speed in problem solving.”

User Training

The participant suggests training the operators to use the system effectively; for example, P8* stated that "Probably a short training or demonstration would be needed to be able to use it properly." This demonstrates that despite the system supporting natural language interaction and being able to explain how to use it, participants still value formal training to improve user experience. Furthermore, P36 stated that "Advantages in the future with generational change with adequate training, no particular difficulties." and P8* also mentioned that once operators get used to the system, it becomes easier, suggesting that with some help, the system was easy to use.

5.4.6. Technical and Operational Issues

Network and Connectivity Issues

The comments indicate significant network and connectivity issues that affected the operations. For instance, P35, and P23 highlight that Wi-Fi and 4G internet within the factory were unreliable. This disrupted communication and data transfer, crucial for smooth user interactions with the assistant as it relied on an internet connection. Additionally, P21 points out firewall issues that hindered access to certain resources. This demonstrates the importance of a reliable and well-managed Wi-Fi network within organizations considering deploying connected systems such as CAs.

Technical Limitations and Troubleshooting

We faced several challenges integrating the system with factory systems. P35 suggests that coordinating with the (3rd party) factory IT was challenging due to their frequent unavailability or preoccupation with other tasks. Furthermore, P21 mentions a technical issue where a PC was inaccessible, highlighting the limitation of coordinating with multiple technical parties to integrate systems.

Dependence on Accurate and Complete Data Input

The comments highlight challenges related to the 'cold start problem': providing benefits and driving engagement if the system's knowledge base is empty. Indeed, during evaluations, P8* indicated that the system needs more data, confirming this issue. Motivating operators to share their knowledge is a considerable challenge, especially initially. Furthermore, once the system has reached a point where it provides perceptible benefits, it will still need to be maintained and kept up-to-date. Factory management also required that the CA include a validation step so that an expert operator can approve all incoming knowledge. Furthermore, P24 emphasizes that the biggest challenge they faced with previous KM systems was keeping them updated due to the significant man-hours required. This suggests a need for more efficient data input methods or automation to ensure the system has accurate and complete data. Additionally, a mechanism is required to remove and update existing knowledge as the machines and products can change over time; P25 states, "Yes, sometimes we have products that change their speed or the machine is modified, allowing faster operation.". This underscores the importance of having an efficient mechanism for keeping a knowledge base up-to-date and approving incoming knowledge.

5.5. Discussion

The domain of Computer-Supported Cooperative Work (CSCW) has evolved significantly over the decades, aligning with technological advancements, a focus on labor relations, and shifts in work practices. Originally, knowledge sharing in CSCW focused on centralizing expert knowledge through the 'repository model' in the late 1980s. This model primarily relied on forming extensive databases that document expert insights. However, this strategy struggled with situational variability—knowledge is often highly contextual and not easily generalized across different scenarios. Furthermore, the requirement for experts to codify their knowledge into repositories imposed a significant authoring burden, detracting from their primary tasks and responsibilities.

Subsequently, the field shifted away from static repositories towards facilitating direct exchanges between individuals, aiming to capture the nuanced and dynamic nature of knowledge. However, scalability remained a challenge, and the effectiveness of these interactions hinged on the availability of both parties, which is not always feasible in a high-paced shift-based production environment.

Most recently, our research aligns with the paradigmatic shift towards cyber-physical systems as intermediaries for knowledge sharing, reflected in CSCW research [31]. These systems, particularly through advancements in AI and live data processing, aim to manage the complexity of real-time knowledge sharing. This leads to critical discussions in CSCW and manufacturing on how to share knowledge from a human to machine and visa versa [40, 226, 227]. Although, still underrepresented, our work contributes to the small but growing body of CSCW work in manufacturing such as [2, 5, 37, 58, 213, 228]. Our work goes beyond prior CSCW work on knowledge sharing in manufacturing such as [5, 27, 58] by learning from the deploying fully functioning CAs at the production line and integrating recent technology advancements, namely LLMs.

Our findings suggest that while (LLM-powered) cognitive assistants effectively bridge gaps left by earlier paradigms, such as scalability and minimizing the authoring burden, they also introduce new dimensions to consider, such as the potential perceptions of surveillance, diminished work fulfillment, and the risks of overreliance on AI. Through this discussion, we situate our work within the broader CSCW knowledge sharing research, highlighting the challenges and lessons learned in the form of design guidelines (**G#**).

5.5.1. Cognitive Assistants Excel at Sharing the Solutions to Emerging Issues

In our study, the CAs were perceived to be effective at sharing knowledge, especially when acquiring knowledge from experienced operators to support novices. Both management and operators recognized that emerging issues at the production line warranted rapid dissemination of solutions that currently did not happen effectively.

As production lines become increasingly complex, digitized, and adaptable, operators spend a considerable amount of time and cognitive resources on error handling, in line with findings of prior work [2]. Concretely, a few hours of downtime could be saved on the production line weekly if solutions to emerging issues are shared effectively. To maximize this benefit, it is crucial to ensure that information can be found quickly and the knowledge base can be easily updated or modified, leading to the following guidelines: **Design CAs to facilitate rapid information retrieval to support timely resolution of issues (G1)** and **Design CAs to enable quick modification and additions to the knowledge base (G2)**. Overall, participants universally agreed on the benefits of sharing solutions to issues such as machine malfunctions or problems caused by human error to resolve the issues quicker.

In contrast to sharing solutions for recurring problems, the topic of sharing knowledge about tuning the machines to achieve higher performance is controversial for some operators. This is especially true for those who are comfortable with operating at lower, stable speeds. These operators were skeptical about the benefits of knowledge sharing on machine fine-tuning as they apply unique strategies that others cannot easily adopt and are highly situated. Machine tuning is a complex and iterative process that is affected by many factors, such as ambient temperature and fluctuating raw material properties, making it more complex to share, aligning with observations by prior CSCW research [5]. On top of this, according to management, some operators attempted to hide their knowledge to maintain their position of power. As such, we observed resistance to sharing valuable knowledge on machine tuning for various reasons.

5

While some experienced operators see no benefit in learning from others concerning machine tuning, this contradicts the management perspective, who observed significant differences in the production speeds attained by different experienced operators and were trying to encourage them to learn from each other. Indeed, other operators still recognized the benefit of sharing some information about machine tuning, such as attainable target production speeds, the accompanying parameters, and some expert tips. Ultimately, in this study, we shifted focus to supporting the operators during issue resolution instead of machine tuning, as the benefits were clear to all, and we wanted to keep the operators interested in the CA. Thus, we would advise to **collaborate with operators to identify the scenarios where sharing knowledge via a CA can result in significant and perceptible support for the operators (G3)**.

5.5.2. Considering the Authoring Burden of Knowledge and the Relevance for other Operators

Regarding the effort of sharing knowledge, several participants brought up the challenge of motivating engagement from operators. While improving accessibility and keeping dialogues short can minimize the perceived effort of using the CA, the perceived benefit of sharing knowledge to help others remains a central barrier to adoption aligning with prior work [27, 58]. Looking at the factors that affect technology acceptance [229], perceived benefit and perceived effort as opposing forces that will impact the attitude toward using the technology. External factors can also influence adoption, such as incentives, cultural norms, and voluntariness to help colleagues. In this study, we observed the strongest motivation to share knowledge among the 'expert' operators who were already involved in training novice operators as part of their job.

Going deeper into the tensions surrounding the balance of providing and receiving knowledge, we assume that more effort invested in eliciting and capturing knowledge will result in a more comprehensive knowledge base that can be used to help others. Understandably, the balance may shift depending on many factors, such as its potential impact, how time-critical it is, if it's something only one expert has discovered, how broadly applicable it is, how situational it is, how readily it can be shared, etc. For example, a solution to a recurring and high-impact problem that only one operator has discovered should be immediately captured and shared, even if it takes considerable effort. Additionally, the act of sharing knowledge is in tension with conducting the primary task of operating the production line, echoing previous CSCW work by Hoffmann *et al.* [5]. It is also important to consider the ephemeral nature of knowledge as the complex system changes and new knowledge is discovered. As such, organizations could consider a prioritization framework to guide operators on how much effort should be invested in sharing their knowledge on a case-by-case basis. This prioritization could be adjusted computationally depending on frequently asked questions that the CA receives as an indicator of relevance or production urgency. **Thus, considering the burden of sharing knowledge on operators, it is important to consider the value of knowledge to prioritize what to share and when (G4)**

Maintaining the temporal relevance and quality of the knowledge base is a key concern from both management and operator perspectives. The managers involved stipulated that all new knowledge entries should be approved by designated 'expert' operators, similar to mechanisms deployed by Hoerner *et al.* [58], providing some level of quality control. In this study, management was enthusiastic about using LLMs to validate the knowledge inputs by checking logic and consistency with existing knowledge. The validation output can be used by the operator to improve the report before submitting it for human approval.

Although not implemented in our work due to the anonymity policies we agreed with operators, other work has made the author of entries visible to others [27], which could support the validation and knowledge application process but also introduce potential bias' concerning the quality of specific operator's knowledge [230]. Additionally, **the validity of entries should be continuously evaluated, for example, by enabling operator feedback or by identifying conflicts between entries and safeguarding against outdated knowledge (G5)**. Overall, several strategies can be employed to maintain a high-quality knowledge base while considering the potential for bias, the burden on knowledge maintenance, and resulting improvements.

5.5.3. Educating the Local Stakeholders is Key

Operators who fundamentally disagree with the benefit of sharing some knowledge may avoid using the CA. Indeed, assuming that the CA can support operators in doing their job better, the operators that are open to using the CA and can effectively interact with it will be at an advantage. For new operators, this means they can become more independent of their human mentors more quickly, and more experienced operators can benefit from the collective knowledge of their colleagues. In contrast, the perceived value of the operators not using the CA may be reduced, for example, because they reject it fundamentally or have difficulties interacting with it. This highlights the importance of **considering accessibility and, where possible, encouraging full operator adoption, for example, by involving them throughout the development process and solving problems that are important to them (G6)**.

During our study, many operators demonstrated a deep pride in the knowledge and expertise they apply in their work. It becomes critical, therefore, to frame technology, such as a cognitive assistant (CA), as a tool that supports rather than replaces their skill and judgment. This notion is in harmony with Shneiderman [101] who advocates for 'humans in the group; computers in the loop', stressing that CAs should enhance rather than substitute human decision-making. **Emphasizing that operators are still responsible for their decisions even with CA support could help them stay critical and informed about the advice they receive from the system (G7)**. Additionally, **communicating the strengths and weaknesses of the CA, such as being aware of potential 'hallucinations' in responses from Large Language Models (LLMs) and the influence of leading questions, is crucial for productive interactions (G8)**.

Yet, similar to prior research, we observed that workers' councils and operators often do not know enough about the technology to engage deeply in discussions concerning new AI systems [228]. On top of this, upon reflecting on our own experience with the workers' councils, we recognize that they were only involved upfront during the project initiation when many details were still fuzzy. In contrast, after the project was approved, they were no longer involved. Considering the innovative and explorative nature of the project, it could have benefited from more continued input from the workers' council to safeguard the operator's interests and ensure they remained well-informed, as demonstrated by Crabtree *et al.* [231] and Zhang *et al.* [232]. We also note that despite our efforts to explain the CA architecture and capabilities in our interactions with operators and posters we put up at the production, we observed that many operators only fully grasped the concept and its benefits once the full system was deployed. This could be attributed to the complexity and novelty of the system we were deploying, which perhaps warrants **additional effort when explaining the concept, for example, through tangible paper prototypes, making the process more accessible to the local stakeholders (G9)**.

5.5.4. Co-Evolving Local Infrastructure and Mitigation Plans for Technical Issues

In addition to the social challenges, deploying CAs at the production line faces many practical and technical challenges that affect the system and our research. The factories in this study used numerous separate digital systems such as for planning, maintenance activities, quality control, productivity, and machine documentation. Many of these systems were considerably outdated and incompatible with each other. Resultingly, considerable effort was needed to connect them to the CA, sometimes requiring the involvement of third-party suppliers. For example, some of the underlying infrastructures, such as databases with API access, needed to be deployed before the research could proceed. As such, considering the effort involved, **relevant stakeholders should discuss whether connecting specific systems to the CA is worth it, considering factors such as added value, reliability, the investment needed, privacy implications, and upcoming system changes (G10)**. As Edwards *et al.* [233] and Martinez-Maldonado *et al.* [225] have highlighted, developing HCI infrastructures that adapt alongside work practices is not just beneficial but essential for the sustainable implementation of technological advancements in any setting.

The poor availability and quality of the data were compounded by poor (wifi and mobile) network stability in the factory, meaning the availability of live data could not be guaranteed. As a result, operators sometimes needed to wait 30 seconds for a response or could not receive the support they requested. Therefore, we advise that **companies should prepare reliable and robust data infrastructures before deploying CAs (G11)**, for example, by providing reliable wifi and investing in consolidating (live) data in centralized databases on modern servers. Acknowledging that high-quality, live data will not always be available, designers and developers of CAs should **design cognitive assistants to be robust to missing or low-quality data (G12)**. For example, mitigation strategies could include proactively asking the operator to provide critical information describing a problem (e.g., pressure levels) if the CA needs it to give reliable advice.

Accessible AI for Support, not Surveillance While production data and issue reports were frequently of low quality or out-of-date, operators were not very motivated to improve them. Several operators mentioned that the current system used for tracking productivity and issues tracking did not accurately measure production speed and sometimes went offline. The operators used this to justify why they did not use it to report issues or pay much heed to the reported production statistics. Then, when a manager tries to investigate why production was low for a shift, the operators claim that the data is inaccurate or does not represent the actual situation. This situation is similar to operator behavior observed by Mörike [37], where operators deliberately entered incomplete data into a digital terminal so that management did not have a complete picture of their work. **Consider that operators may resist systems that provide managers with more accurate and timely information on production processes (G13)**. Doing so gives the operators more control which can make their work more meaningful [234, 235] Thus, understanding the origins of operator resistance to system changes is vital for designing CSCW and HCI solutions that will be adopted and used effectively.

Although most operators expressed concerns about privacy, particularly regarding surveillance by managers and being unfairly evaluated on work performance, the majority agreed to allow sensitive data collection, such as their location. The main requirement was that access was restricted, particularly from management, highlighting the necessity that the benefits of data collection should be clearly beneficial to operators and not just for surveillance, echoing findings of Das Swain *et al.* [236] that found information workers perceived tracking as pure surveillance when they were not presented any consumable insights. Surprisingly, the operators we interviewed went as far as to express favoring a more intrusive tracking system that could enhance the system's effectiveness over half-measures that did not benefit them. Thus, **any implementation of tracking technologies must carefully balance utility and privacy while including operators in the decision process (G14)**.

A major challenge when involving operators in privacy discussions is catering to widely different perspectives. In our study, the cognitive assistant has context awareness through an operator tracking system. The tracking system consisted of multiple stereoscopic cameras and a computer vision model to create an anonymous 18-point skeleton of operators along the production line. The tracking data was not personally identifiable, and factory employees could not access it. The workers' council and all the operators we spoke to approved the tracking system. Furthermore, all operators were informed about the tracking system via announcements and posters along the production line. Despite these efforts, the tracking system was sabotaged by an operator. Despite pressure from factory management to fix the tracking system and continue with the research, we interpreted the operator's actions as an implicit rejection of the tracking system and disabled it. Thus, beyond respecting the legal side of privacy, organizations will need to consider when to force the implementation of human tracking for CAs, especially if the majority of the workforce would prefer the additional utility that context awareness could provide, such as automatically capturing knowledge [230] or suggesting responses.

Beyond context-aware suggested responses, several other factors can impact the accessibility of interactions with the CA at the production line. These factors include familiarity with CAs, difficulty reading text on a smartphone screen, physical access to the smartphone with the CA, noise that may impact the accuracy of speech-to-text and audibility of text-to-speech, (local) accents or jargon affecting speech-to-text and NLP, and existing knowledge on the abilities and limitations of AI technology, such as LLMs. Indeed, aligning the design of the CA with the needs of operators is a crucial area of research [237]. The challenges of handling noise, accents, and local jargon when using speech-to-text remains an open point but beyond the scope of our work. In this study, we observed that providing (context-aware) suggested responses for the operators to click on instead of type responses was effective at improving accessibility. Furthermore, using advanced NLP technologies, such as LLMs, made the CA easier to use as it became more capable in understanding the operator's queries. The final version of the CA we deployed was also accessible in computer browsers instead of only through a smartphone app, improving accessibility for some of the older operators who had trouble using the smartphone. These integrations underline the **importance of designing technology that is adapted to real-world factory contexts and the capabilities of the operators (G15)**.

5.5.5. IMPLICATIONS FOR RESEARCHING AND DEPLOYING COGNITIVE ASSISTANTS IN FACTORIES

Navigating the tensions between operators and management regarding AI-mediated knowledge sharing is key to the success of CAs in factories. As the management initiates the decision to bring in new technology like CAs, the operators can perceive this as a means of control and surveillance. As discussed by prior studies, the early and continued involvement of operators and workers' councils can facilitate the adoption of new technology but did not go into the specifics on how [213, 238]. The technology investigated in this study, (LLM-powered) cognitive assistants that support knowledge sharing, pose some unique challenges that warrant extra attention, such as the difficulty imagining the impact on work, the authoring burden on (expert) operators, focusing on knowledge that operators understand the benefits of dissemination, understanding the limitations of LLMs and the possibility of feeling surveilled by management. Unlike prior work that was forced to have managers mediate all interactions with operators [213], we were able to involve them separately in most cases. Thus, **we could maximize the openness and directness of the responses we received by positioning ourselves as external researchers (i.e., not 'agents' of the management), building a relationship with participants over time, talking to operators without management present whenever possible, and maintaining the anonymity of participants (G16)**.

From our work, we recognize that the most **crucial topic to involve operators is identifying what knowledge would be beneficial to share and how it can be expressed and retrieved (G17)**. Tackling this together requires clear communication and education about what the technology can and cannot do. In this study, we witnessed the importance of respecting the operator's expertise and autonomy, making sure that operators feel their contributions through the cognitive assistant are valued, not just as data for the system but as important knowledge that improves the workplace. For example, by focusing on sharing knowledge about resolving concrete machine issues as opposed to the less tangible machine tuning knowledge, more operators perceived the CA to be beneficial. This is especially important as many operators were apprehensive of introducing (yet) another tool and the potential consequences thereof, similar to observations made by [238]. By **ensuring everyone understands and sees the benefits of cognitive assistants, factories can reduce resistance and increase both productivity and job satisfaction (G18)**.

Regarding social interactions, while AI allows for quick access to information, it could also lead to fewer face-to-face interactions, affecting the social environment. Additionally, increasing dependence on AI means operators must develop new skills, particularly in using advanced digital tools, instead of relying solely on experiential knowledge. This shift could also affect operators' well-being, balancing between the satisfaction from autonomously or collaboratively solving complex problems and the ease of AI help as previously discussed by Wurhofer *et al.* [2]. We observed that the operators were proud of their expertise, and introducing the CA to capture and share this might affect the satisfaction they get from work.

While using a CA for knowledge sharing promises to enhance operational efficiency, **it is vital to balance using the AI tool for all information needs versus the irreplaceable value of human knowledge sharing, if available (G19)**. In fact, several operators preferred learning from human operators, similar to findings in prior work [5]. Yet, a key benefit of the tool is that it could help with problems when a knowledgeable human colleague is unavailable or unable to help.

The long-term effects of using AI for sharing knowledge in factories are likely complex, covering operational, psychological, and social aspects. Using AI for this purpose might improve operational efficiency and make the knowledge management system within organizations more dynamic and responsive. This could change how organizations are structured, making operators more independent of other (expert) operators. However, there may be fewer opportunities to share more subtle, tacit knowledge that is more challenging to share through technology, emphasizing the need for research in these areas.

Reliance on AI for decisions may also lead to poorer decision-making in the long-term, aligning with concerns raised by Buçinca *et al.* [239]. Furthermore, while an expert operator might intuitively be able to spot LLM 'hallucinations', novice operators may not have sufficient experience to do so. Even if accurate, in the long term, overreliance on AI might weaken critical thinking and problem-solving abilities among operators. These challenges highlight the importance of longitudinal studies at factories. However, based on our experience navigating the safety and productivity concerns of factories during this study, this will be challenging to conduct. Aligning with prior work in this area, this highlights the need for methods that will allow the study of long-term impact in a way that is acceptable to industrial partners [213].

5.5.6. Limitations

In conducting this study, we faced numerous challenges tied to fieldwork in operational settings where financial, health, and safety concerns are paramount. Consequently, the tested prototypes were never utilized for extended periods during actual work activities; instead, we relied on a series of brief user tests, each lasting about 30 minutes. Despite the limitations posed by these short sessions, we are confident in the ecological validity of our findings given the high technology readiness levels (TRLs 7-8) of the prototypes, which were consistently integrated and tested within the real-world factory environment.

While our rapid prototyping approach facilitated immediate feedback and iterative improvements, a more prolonged deployment would likely uncover deeper insights into how users gradually adapt to and integrate cognitive assistants in their daily workflows. Such extended usage could provide a clearer picture of the shifts in user acceptance, the evolving effectiveness of the system, and the potential unforeseen impacts on workplace dynamics and operator autonomy.

5.6. Conclusion

Reflecting on the work presented in this chapter, it represents an evolving, longitudinal study. Over the course of three years, our understanding of CAs and their integration in factory settings to enhance knowledge sharing among operators has deepened. Each study conducted was part of a continuous learning process, where early findings and technological breakthroughs affected subsequent prototypes and research questions. The findings should be viewed as part of a developmental trajectory rather than isolated outcomes. Furthermore, our own relationship with the factory personnel evolved over the years and may have been shaped by increasing familiarity and trust, which could have led to more candid feedback over time.

This study highlights the complexities and potential of integrating cognitive assistants (CAs) in factories to support knowledge sharing among operators, emphasizing the necessity of addressing both technological efficacy and organizational dynamics for successful implementation. Efficient knowledge sharing through CAs can significantly expedite problem-solving and help novices become more autonomous, yet hinges on the systems' abilities and overcoming adoption challenges. To promote the adoption of CAs, surveillance concerns should be addressed transparently, preferably in a setting where operators are comfortable sharing their perspectives. Involving all stakeholders, including workers' councils and operators, in developing CAs is crucial for trust and acceptance, where stakeholders are sufficiently informed and educated to engage in discussions.

6

Discussion and conclusion

In this dissertation, we explored the design, engineering, and deployment of cognitive assistants (CA) in factories, focusing on improving Knowledge Sharing (KS) among operators. This work explored both the potential benefits and the significant challenges of deploying advanced AI technologies in such complex environments.

Our studies have revealed important insights into design opportunities, user interaction, privacy concerns, and the balance needed between technology, social, organizational, and human factors for successful CA implementation. Our findings highlight the promising role of CAs in improving efficiency, learning, and problem-solving on the production floor. However, they also emphasize the need for designs that center on human needs, organizational context, robust NLP, and ethical considerations that consider the perspectives of all stakeholders involved. Looking ahead, future work should focus on investigating the long-term impact of CAs for KS and exploring how to refine CAs for KS further. Broader research should investigate how our findings apply in other settings and frameworks to inform future research and design. These topics are discussed across four main sections of this chapter, namely: 6.1 Answers to the research questions, 6.2 Implications, 6.3 Reflections on the approach, and 6.4 Future work.

6.1. Research Questions Answered

In the following sections, we answer the four main research questions tackled in this dissertation (see Section 1.5.3).

6.1.1. RQ1: Initial Design, Opportunities, and Challenges

The introduction of CAs in factory settings aims to improve KS using conversational AI. Yet, existing research has not investigated their design requirements, usage scenarios, and risks. This led to the formulation of our first research question: **RQ1: What are the opportunities and design challenges when deploying cognitive assistants to facilitate knowledge sharing among factory operators?** We engaged in ethnographic research across multiple factories, which informed the design of a CA, potential interaction scenarios, and the identification of challenges and risks [102].

The proposed system design for the CA integrates AI capabilities to facilitate KS in factory settings. Central to this system is a **knowledge graph**, which is dynamically updated with operator knowledge situated in context. In addition, the CA is **context aware** by integrating live data from the production line and an anonymous operator tracking system. The anonymous operator tracking system used stereographic cameras and a computer vision model to create 18 XYZ-coordinates representing the operator's body as they move around the production line. This enables the assistant to provide context-aware support depending on which machine the operator was (recently) standing next to (see Section 2.3 for more details). Operators interact with the CA through a **mobile app** designed for various input and output modes, including voice, text, touch, and visual prompts. This multi-modal approach makes the system accessible and convenient for all users, ensuring quick and relevant assistance. The app aims to reduce interaction time while providing contextual, actionable advice, making it easier for operators to get help or share information. By **proactively engaging operators based on their location, current activity, and machine statuses**, the assistant not only helps solve immediate problems but also promotes ongoing learning and improvement. To ensure a successful design and deployment, we identified several opportunities and challenges, which are outlined below:

Design Opportunities and Challenges

6

Context aware knowledge capture. The ability to track operators and process machine data gives the assistant context awareness, a key enabler for interacting with factory knowledge [144, 240]. These data streams can reduce interaction time by automatically prefilling information (e.g., when a problem occurs, the CA can automatically retrieve error codes and machine names and settings), identifying best practices by observing operator interactions with machines, and identifying opportune moments to interact with the operators.

Balancing efficient problem-solving and learning. Critical engagement could also be an issue when the assistant advises the operator. Operators who become overly reliant on the CA and knowledge shared by others may lose their own critical thinking and problem-solving abilities. A key challenge, therefore, is balancing quick assistance with operator learning and active decision-making. Ensuring the CA does not encourage an over-reliance among operators on the CA for situational awareness, problem-solving, and decision-making requires a design that promotes critical engagement, for example, by promoting reflection on the decision [241] and providing appropriate explanations [242]. That being said, introducing a reliable CA for information retrieval and memory storage could allow operators to forget certain routines to make space for other knowledge [243].

Balancing interaction speed with utility. Designing efficient user interactions to capture knowledge in a noisy, jargon-heavy, fast-paced factory environment is challenging. Challenges include ensuring that voice commands are accurately recognized in noisy areas and that all the nuances and technical jargon are accurately processed [23]. While quick user interactions could boost KS, capturing and codifying tacit knowledge — especially the inexpressible experiences and insights of experienced operators — poses a significant challenge. Balancing the system's information needs with the need for intuitive and straightforward user interactions is a critical challenge. Ensuring the completeness and accuracy of this knowledge requires innovative solutions that facilitate efficient knowledge elicitation methods without oversimplifying the richness and contextual situation of human expertise [244]. However, achieving this level of adaptability demands advanced NLP models that can reliably handle the nuanced, contextual, and social aspects of real-world factory operations [75, 245].

Safety and ethics. Prioritizing operator safety and ethical considerations, particularly data privacy and knowledge attribution, introduces complex challenges. Addressing the ethical challenges requires transparent policies and secure data management practices that protect operator rights and company knowledge while minimizing the impact on the functionality and efficacy of the assistant. For example, while technological solutions such as operator tracking (see the anonymous operator tracking system presented in Section 2.2) might improve the contextual awareness of the systems, it introduces several ethical concerns related to the operator's privacy [246]. Moreover, inherent risks associated with introducing new AI tools into demanding, high-risk work environments require careful consideration. The same holds for conducting research in the environment, as any distraction to the operator may result in a loss of situational awareness or heightened cognitive load, leading to poor decision-making.

Interaction Scenarios

Based on our context analysis of factory operations and system capabilities, we identified several usage scenarios where the CA can be optimally used. During **production line downtime**, the system could alert operators about the production line status, recommend relevant training, and provide feedback on their contributions, tailored to avoid repetitive training and identify when refreshers are needed. During **production line reconfiguration**, the assistant is intended to offer contextual recommendations, guiding (novice) operators with best practices and answering queries based on their proximity to machines and activity. In addressing **operational issues**, the assistant is envisioned to suggest investigative steps, assist in identifying root causes using the 5-whys technique, employ similarity models for unfamiliar problems, and leverage machine learning to preemptively notify operators about potential issues, thus aiding in the rapid resolution and effective documentation of problems. These implementations expand the assistant's knowledge base while optimizing operational workflows.

Aligning with related work on using assistants to support factory operators, it is evident that deploying CAs in factory settings stands to benefit operators but also faces multifaceted challenges [60, 144, 147, 245, 247]. Addressing the technical challenges while considering human, organizational, and ethical factors, will be key to realizing the potential of CAs to improve KS in manufacturing.

6.1.2. RQ2: Effects of Modality, Training, Context Expertise on UX, Usability and Interaction Efficiency

Building on the key opportunities and challenges identified in Chapter 2, the second research question focused on deepening our understanding of how decisions regarding CA design and evaluation affect user experience: **RQ2: How do modality, user training, and**

context experience affect the UX, usability, and interaction efficiency of CAs for KS in factories? To investigate this question, we developed a functional CA and simulated machine interfaces and conducted an empirical study. The study included participants from various backgrounds, such as factory operators, students, and HCI researchers. Additionally, through the qualitative analysis of user feedback, we identified key areas for improvement and established lessons learned [103]. The insights are discussed below, arranged under four topics: user insights, modality, training and context experience.

User Insights for Design Optimization

The feedback on the prototype unveils crucial considerations to inform the design of CAs. Efficiency when using the assistant was paramount—users wanted to complete their tasks quickly—thus, anything that prevented them from doing so was a burden, as prior work in other contexts has also shown [248]. For instance, users' struggle with frequent conversation breakdowns, and ineffective repair strategies—reflecting the unreliability of traditional NLP—indicating a need for more robust NLP systems before operators will adopt the system. Understanding user intent from shorter (incomplete) sentences that contain a few keywords, as suggested by participants, might alleviate some difficulties users face, enabling a more flexible and efficient interaction by allowing single-word commands for frequently used features. This echoes prior findings showing that users tended to use shorter sentences when interacting with chatbots [249]. Indeed, as users become more familiar with the NLP capability and limitations, they will likely adjust their phrasing to ensure more successful interaction. Alternative strategies include more extensive use of buttons (suggested phrases), autocomplete, or training [250]. Much of our feedback highlights the shortcomings of intent-based NLP, which does not handle nuanced language or divergent phrasing well [158, 248].

Modality: Smartphone versus Laptop

At the production line, a computer is available for operators to use. We assessed if interacting with the CA on the computer versus a smartphone impacted the UX, workload, or interaction efficiency. The lack of significant differences between smartphone and laptop modalities in task performance metrics suggests an opportunity to prioritize mobile-friendly CAs in factory settings. Given the portability of smartphones, designing CAs with mobile-first UX could not only cater to the existing habits of operators (many of whom may already rely on smartphones for various tasks) but also ensure that assistance is readily available at any location on the production floor. Although others advocate for using tablets in factories for user interactions [251], we believe that the portability and familiarity of smartphones is a key benefit [252], as operators can easily carry them around in their pockets as they move around the production line.

Training: A Double-Edged Sword

Interestingly, the study reveals that prior training to use the system does not necessarily correlate with improved task performance or user satisfaction. Training may have heightened expectations or increased cognitive workload, as users possibly tried to recall training instructions during interaction. Indeed, prior work has shown that the timing of providing training to users can have a significant effect on how it is perceived, even if it helps complete tasks quicker, emphasizing the strong psychological factors involved [250]. Ideally, the need for extensive training should be avoided by designing systems that are intuitive and self-explanatory. This can be particularly relevant in fast-paced manufacturing environments where time is a critical factor, and users may not adopt a system that requires a lot of training to use effectively [253].

CONTEXT EXPERTISE: LAYMEN ARE MORE CRITICAL OF THE SYSTEM

The difference in usability scores between operators and laymen highlights the influence of context expertise on user experience and perceived usability, namely, that factory operators were generally more positive than laymen when assessing the assistant. Operators' familiarity with manufacturing environments likely set a baseline for their expectations and perceptions of utility, different from that of laymen. This discrepancy underscores the importance of involving end-users in the design process—ensuring that CAs meet the specific needs and expectations of their intended user base.

Overall, this study demonstrated the stark differences in study results when testing systems with end-users with context expertise versus laymen. Furthermore, it emphasized the importance of efficient and reliable interactions and the shortcomings of traditional, intent-based NLP.

6.1.3. RQ3: HARNESSING LARGE LANGUAGE MODELS FOR KNOWLEDGE SHARING WITH CONVERSATIONAL AI

Building on the knowledge presented in prior chapters on CAs' designs, we investigated how state-of-the-art NLP technologies can improve their value further. This section addresses **RQ3: What are the implications of using LLM-based conversational AI for knowledge sharing among factory operators?** To explore this, we designed and evaluated an LLM-based CA system in a manufacturing setting and conducted a benchmark study comparing various open-source and proprietary LLMs.

SYSTEM DESIGN AND EVALUATION

The CA system leverages Large Language Models (LLMs) to improve KS in manufacturing environments, addressing the traditional challenge of accessing and sharing high-volume, complex operational knowledge. The system employs Retrieval Augmented Generation (RAG) to efficiently respond to operator queries by retrieving and processing information from extensive factory manuals and issue reports. Additionally, it supports knowledge authoring through a dynamic issue reporting form using the 5-whys method and an LLM-powered knowledge validation step. This approach ensures high-quality knowledge authoring while minimizing the burden on contributors and validators.

A field study conducted with nine participants from a detergent factory, including both managers and operators, demonstrated the system's usability and effectiveness. The study highlighted the potential of LLM tools to significantly improve technology-supported KS in manufacturing environments.

BENCHMARKING STUDY

The benchmarking study compared the performance of various proprietary and open-source LLMs in responding to manufacturing-specific queries. The assessment focused on the models' ability to provide factual, complete, and concise answers based on factory operational documents and issue reports. GPT-4 emerged as the leading model, showcasing superior accuracy and minimal hallucination in responses. Open-source models like StableBeluga2 and Mixtral 8x7B also performed well, indicating significant progress in open-source LLMs, which are getting closer to proprietary models in domain-specific contexts such as manufacturing.

The research revealed several key findings. First, LLM-powered systems can increase information accessibility by extracting information from poorly structured documents, enabling the assistant to harness a wider set of existing documents and process unstructured human inputs, such as voice or chat reports from operators, corroborating findings by Kiangala and Wang [59]. Despite this, some operators expressed a preference for seeking help from human operators if available, deeming it more efficient and accurate. However, knowledgeable human operators are often unavailable during shifts, making the LLM-powered assistant a valuable alternative. Additionally, the performance gap between open-source and proprietary LLMs is rapidly closing, suggesting that open-source, locally hosted LLMs may soon be or already are viable options for manufacturing environments where data security and control are critical, as indicated by general LLM benchmarks [22].

The research demonstrates that LLM-powered CAs can significantly improve KS and operational efficiency in manufacturing settings. By effectively retrieving and processing complex operational knowledge, these systems address a core need in the industry. The benchmarking study provides insights into selecting appropriate LLM tools, considering performance, privacy, and customization requirements, thereby enhancing KS practices within manufacturing environments.

6.1.4. RQ4: FACTORY OPERATOR PERSPECTIVES ON THE SOCIO-TECHNICAL CHALLENGES AND PERCEIVED IMPACT OF COGNITIVE ASSISTANTS

Technological progress may facilitate knowledge-sharing in factories, yet the successful deployment of CAs depends on understanding the complex interplay between technology and human factors, as well as the viewpoints of all stakeholders involved. In light of this, our investigation centered on the following question: **RQ4: What are factory operators' and management perceptions of the impact and socio-technical risks and challenges of using conversational AI for knowledge sharing?** To answer this question, we conducted a thematic analysis based on the end-user feedback from deploying these assistants at two factories for over two years [108].

COGNITIVE ASSISTANTS REDUCE PROBLEM SOLVING TIME

Factory operators and management perceive the impact of using CAs as multifaceted, including both opportunities for improved efficiency and socio-technical risks and challenges. Operators and management acknowledge the potential of CAs to swiftly retrieve information and assist in problem-solving. This underlines the potential of these tools in facilitating effective KS and possibly improving the autonomy of novice operators. For expert operators, the tool serves as a personal memory augmentation and an efficient way to disseminate solutions to emerging issues. Participants acknowledge that operators waste hours every week solving previously solved issues that the system could help prevent, representing situations where the CA can help save time.

Socio-Technical Implications and Concerns

Beyond the upfront benefits, concerns are raised about the socio-technical implications of integrating CAs into factory operations. There are apprehensions regarding the systems' ability to handle complex, nuanced knowledge and the risk of creating a dependency on technology that may not always be reliable or available. Concerns of operators becoming over-reliant on AI for decision-making match the findings of prior research. This risk has motivated recent work on mitigating overreliance on AI [239, 254, 255].

The need for these systems to be user-friendly and accessible for all operators, irrespective of their tech-savviness, is stressed. We also observed that while end-users were generally positive, a minority was pessimistic about the potential benefits, and at least one end-user outright sabotaged the system, which we believe was related to a mistrust regarding whether managers could access the raw camera feeds. While we attempted to balance the needs of the organization and individual operators, these examples demonstrate the fragility of the process. Perhaps, a more incremental, "muddling through" approach [256], would have been successful in user adoption. On the other hand, the clear rejection of at least one operator may also point to underlying trust issues toward the management that the CA may exacerbate.

Barriers to Adopting Cognitive Assistants

The adoption and change management processes underscore the divide in perspectives between the operators and management, revealing concerns over privacy, data handling, and empowerment. Resistance to change, fueled by previous bad experiences and anxiety over new technology integration and privacy, shows the social and organizational rifts that CAs might exacerbate if not addressed thoughtfully. The operator activity tracking functionality was the most contentious functionality related to privacy and was covertly sabotaged by an operator. Interestingly, this objection was not raised during the initial talks with the operators and the worker's council, perhaps because the operators did not want to draw too much attention to themselves. As such, only deploying the system at the factory premises led to this insight, demonstrating the importance of iterative testing and evaluation in the field.

Security concerns about misinformation and the ethical dimensions of operator autonomy further complicate the adoption of CAs. These challenges clearly demonstrate the importance of regularly reevaluating privacy-related aspects, benefits, and risks. Simultaneously, technical issues such as network connectivity and data accuracy emphasize the infrastructural and logistical hurdles, key barriers to smart manufacturing in developed countries [257]. Ultimately, balancing the practical potential with the human-centered needs of the workspace emerges as a concern for both operators and management, highlighting the need for a human-centered approach when deploying CAs in factory settings.

6.2. IMPLICATIONS

While using a CA for KS promises to increase operational efficiency, it is vital to balance against the irreplaceable value of human KS, if available. The tool will have the most significant impact when operators face a recurring issue that someone else has already faced and documented but is unavailable for direct consultation. Perhaps the most important distinction to existing systems for acquiring knowledge in factories [58, 75, 244] is that they require human analysis of data or knowledge elicitation. In contrast, the CA systems described in the dissertation capture this knowledge automatically through interactions with operators. Recent advances in NLP made the automatic capture and processing of knowledge more feasible [258], attaining sufficient reliability and accuracy to process unstructured natural language [59, 259] and interact reliably with operators on the work floor. Practitioners developing these systems must focus on creating interfaces and communication protocols that maximize the benefits of LLM-driven interactions while mitigating the risks, such as "hallucinations". In the following sections, we discuss the implications and guidelines for using CA in factories.

COGNITIVE ASSISTANTS EXCEL AT SHARING SOLUTIONS TO EMERGING ISSUES

In our studies, the **CAs were perceived to be effective at sharing knowledge, especially when acquiring problem solving knowledge from experienced operators to support novices**. Both management and operators recognized that emerging issues at the production line warranted rapid dissemination of solutions that currently did not happen effectively. As production lines become increasingly complex, digitized, and adaptable, operators spend a considerable amount of time and cognitive resources on error handling, in line with findings of prior work [2]. Concretely, a few hours of downtime could be saved on the production line weekly if solutions to emerging issues are shared effectively. To maximize this benefit, it is important that information can be found quickly and the knowledge base can be easily updated.

MAINTAINING HIGH KNOWLEDGE QUALITY

Beyond being able to update the knowledge base, **mechanisms should be in place to safeguard the quality of the knowledge**. Throughout this dissertation, we explored different mechanisms such as ratings based on feedback, ranking suggestions based on performance, using LLMs to check any new knowledge for possible inconsistencies or conflicts, and requiring a human expert to approve every new addition. Considering that many of these mechanisms result in additional work, organizations will need to balance the need for strict quality control with additional labor costs. Although we opted to anonymize knowledge contributions for privacy reasons, organizations could consider attributing knowledge to the contributors by name as done in work by Hannola *et al.* [27]. As factory operators might perceived some of their colleagues to be more knowledgeable or trustworthy than others, they can use this information in their decision making process. This may also affect their sense of community or offer incentives to share. Thus, organizations must find mechanisms for quality control of knowledge that match their goals and context.

Identifying Valuable and Sharable Knowledge

From our work, we recognize that it is vital to involve operators when **identifying what knowledge would be beneficial to share and how it can be expressed and retrieved.** Tackling this together requires clear communication and education about what the technology can and cannot do. In this dissertation, we witnessed the importance of respecting the operator's expertise and autonomy, creating an environment where operators feel their contributions through the CA are valued, not just as data for the system but as important knowledge that improves the workplace. For example, by focusing on sharing knowledge about resolving concrete machine issues as opposed to the less tangible machine tuning knowledge, more operators perceived the CA to be beneficial. This is especially important as many operators were apprehensive of introducing yet another KS tool and the potential consequences thereof, similar to observations made by [238].

Be Wary of Hallucinations when Integrating Large Language Models

The adoption of LLMs is not without its risks. The potential for generating incorrect information or "hallucinations" necessitates comprehensive risk management strategies. **Practitioners must develop and implement safety nets and verification processes to ensure the reliability and accuracy of the information generated by LLMs.** Moreover, as technology evolves, keeping up-to-date with new techniques to reduce hallucinations and improve reasoning capabilities will be key. For example, the following techniques and insights have been shown to improve the quality of responses: chain-of-thought prompting [96] and its derivatives such as graph of thought [260], cumulative reasoning [261], and Tree of thoughts [262]; meta prompting [263]; self-discovery [264]; and algorithms for mitigating "lost in the middle" issues [265] for long contexts such as LongLLMLingua [266]. For policymakers and legislators, this aspect underlines the need for regulations that ensure these technologies are developed and used with mechanisms that maximize accuracy and reliability.

Tensions between Management and Factory Operators

Navigating the tensions between operators and management regarding (AI-mediated) KS is key to the success of CAs in factories. Considering that management initiates the introduction of new technology like CAs, the operators can perceive this as a means of control and surveillance. The perception of control appears to stem from feeling forced into sharing knowledge, which factory operators felt they were already doing well enough. Meanwhile, the perception of surveillance stems from the CA system's ability to collect information about their location, actions, productivity, or knowledge, providing managers with more visibility of their work. Despite these initial concerns, we observed that over time, and given sufficient opportunities to provide input in the design process, operators began to perceive the CAs as supportive tools instead. This matches conclusions drawn in previous work where the early and continued involvement of operators and worker's council was seen to facilitate the adoption of new technology [213, 238].

That being said, this shift was not universal, with some operators remaining distrustful and uninterested in using CAs. Thus, factory operators who never adopt the system may end up at a disadvantage compared to their peers.

Reflecting on how factory operators spoke about existing issue reporting and performance tracking systems, we observed that they often challenge the accuracy and reliability of the systems. By undermining the system's accuracy, the operators could attribute their lower production performance figures to inaccuracies in the measuring system. Similar observations have been made in prior work when digital reporting systems were introduced, where factory operators purposely submitted incomplete entries to obscure their actual performance from management [37]. In contrast, the managers we spoke to in the context of this dissertation stood by the accuracy of the production metrics and reliability of the issue reporting systems. Thus, it is **important to maximize the perceived accuracy and trustworthiness of the CA while also making clear that the CA is not intended to be a tool for surveillance or performance evaluation**. That being said, it might be necessary to mask operator identities, timestamps, or other personally identifiable data to convince operators that it cannot be misused.

Long-Term Human and Organizational Implications of AI for Knowledge Sharing

6

The long-term effects of using AI for sharing knowledge in factories are complex, covering technological, psychological, and social aspects. On one side, using AI for this purpose might improve operational efficiency and make the knowledge management system within organizations more dynamic and responsive. This could change how organizations are structured, making operators more independent of other humans. However, this change also risks too much reliance on AI for decisions, which can lead to poorer decision-making [239]. In the long term, overreliance on AI might weaken critical thinking and problem-solving abilities among operators. These (potential) consequences highlight key areas for future research [267]. However, based on our experience navigating the safety and productivity concerns of factories during this study, a long-term study will be challenging to conduct. Aligning with prior work in this area, this highlights the **need for methods that will allow the study of long-term impact in a way that is acceptable to industrial partners** [213].

Impact on Social Interactions and Factory Operator Well-Being

Regarding social interactions, while AI allows for quick access to information, it could also lead to fewer face-to-face interactions and learning from human contact, affecting the social environment and group intelligence at work. Additionally, increasing dependence on AI means operators must develop new skills, particularly in using advanced digital tools, instead of relying solely on manual or experiential knowledge. This shift could also affect operators' well-being, striking a balance between the satisfaction from solving complex problems and the ease of AI help, matching observations made by Wurhofer *et al.* [2]. We observed that the **operators were proud of their expertise, and introducing the CA to capture and share this might affect the satisfaction they get from work.** This may be compounded further if the CA reduces the need for factory operators to work together to solve problems.

Finally, all these factors together play a role in the long-term success of companies; organizations that strategically apply AI for KS could gain competitive edges through faster problem solving and increased productivity. Yet, achieving this requires designing AI systems that cognitively support humans without removing the human and cultural elements vital for creating a learning-focused organizational culture.

Confidentiality Implications of using Large Language Models with Organizational Knowledge

Confidentiality considerations surrounding data privacy and security become critical with integrating data-driven AI technologies [268, 269]. Relying on considerable computing power and third-party services introduces significant data security concerns for the companies. Practitioners developing LLM systems must consider secure data management practices. **Open-sourcing of LLM technologies by developers—pioneered by Meta with Llama—is a promising step towards democratizing these technologies, potentially allowing for local hosting that maintains high performance while alleviating some security concerns.** The gap between proprietory LLMs such as GPT-4-turbo[1] and Claude 3 Opus[2], Gemini 1.0 Ultra[3], and open-source alternatives such as Mixtral 8x7B[4] and Llama 3[5] has closed significantly since the launch of the first major open-source LLM, Llama [204]. However, the hosting efforts and investments required, mean organizations will likely continue relying on large, established cloud service providers such as Azure[6] or AWS[7]. This reality stresses the importance of policymakers and legislators creating a regulatory environment that encourages secure data practices.

6

6.3. Reflections on the research approach

Engaging end-users in the design and development processes of CAs is critical. This collaborative approach ensures that the tools address **real-world needs and challenges**, facilitating a sense of ownership and acceptance among users. The requirements and concepts initially put forward by management did not match the needs of shop floor operators and were met with disdain. However, we observed **increasing enthusiasm for the tool as the project progressed**. From disinterest initially, to enthusiasm at later stages. We believe a large part of this can be attributed to allowing the operators to voice their perspectives and ensuring that the tool would benefit them while also meeting organizational needs. Furthermore, the benefits were not directly obvious to all operators, requiring in-depth discussion to reach sufficient understanding. We also recognize that it can be difficult to express the benefits of a novel tool being actively developed and

[1] `https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo`—last accessed November 20, 2024
[2] `https://www.anthropic.com/news/claude-3-family`—last accessed November 20, 2024
[3] `https://gemini.google.com/advanced`—last accessed November 20, 2024
[4] `https://mistral.ai/news/mixtral-of-experts/`—last accessed November 20, 2024
[5] `https://llama.meta.com/llama3/`—last accessed November 20, 2024
[6] `https://azure.microsoft.com/en-us/products/ai-services/openai-service`—last accessed November 20, 2024
[7] `https://aws.amazon.com/machine-learning/inferentia/`—last accessed today

adapted as the project progressed. At the project's conclusion after three years, many operators appreciated the tool's benefits.

Our approach incorporated several methods, including **research-through-design, ethnographic research, co-design sessions, and empirical studies**. The **foundation of this project on ethnographic research** has provided valuable context and understanding of the factory environment, which has directly informed the design and development of the CA. The **research-through-design approach was instrumental in developing CAs, allowing researchers to explore novel interactions and interfaces**. It facilitated a hands-on understanding of the problem space, enabling the iterative refinement of the CA prototypes based on real-world usage and feedback.

The prototypes, especially the later ones, were real systems integrated into factory environments, which stemmed from the requirements of the participating factories. While the combination of fully functional prototypes in a factory environment provided a high ecological validity for our research, the necessity of creating advanced, innovative prototypes sometimes distracted us from the human and organizational factors. Thus, on reflection, there were periods when the development work became very prominent and focused too much on the artifact being designed (the CA) and **less on the broader system and organizational context** in which it operates, potentially overlooking human factors critical to successful implementation. This points to the importance of frequent moments of reflection to take a step back from the low-level design details.

Our access to factory stakeholders was significantly restricted by factory management to minimize the impact on production, possibly affecting the depth and candidness of the responses. While our ethnographic work, prototype evaluations, and co-design sessions aimed to include a range of stakeholders, it was most challenging to recruit operators as they were only available for short periods during their shifts at the production line. This could potentially skew the design toward the needs and perspectives of senior operators and managers. Furthermore, the shorter evaluations with the operators at the production line **focused primarily on immediate usability concerns and imagined long-term benefits**, with less emphasis on long-term impact and user engagement. As a result, the **assessments of the final designs were largely qualitative**, making it challenging to measure the impact quantitatively or compare it against other solutions.

That being said, unlike prior work that was forced to have managers mediate all interactions with operators [213], we were able to involve them separately in most cases. Thus, we could maximize the openness and directness of the responses we received by positioning ourselves as external researchers (i.e., not 'agents' of the management), building a relationship with participants over time, talking to operators without management present whenever possible, and maintaining the anonymity of participants.

6.4. FUTURE WORK

In discussing future work, we address research that directly continues from this dissertation and broader future work.

6.4.1. IMMEDIATE EXTENSIONS

REFINEMENT OF LARGE LANGUAGE MODEL-POWERED COGNITIVE ASSISTANTS

Refinement of LLM-powered CAs within the factory context would entail enhancing the accuracy, reliability, scalability, and context understanding. Efforts could include refining the retrieval augmented generation pipeline, experimenting with a modular approach, and more in-depth optimization of the (hyper)parameters. Furthermore, we suggest investigating interactions to support even more efficient knowledge elicitation, validation, and maintenance, for example, by enabling operators to record short (audio)visual notes and delete and/or replace entries in the knowledge base.

LONGITUDINAL IMPACT STUDIES

Long-term studies are needed to assess the enduring impacts of CAs on a human, team, and organizational level. On a human level, key positive impacts could include skill development and learning, critical thinking, and well-being. On the other hand, we might also observe losses in skills, such as memory, as operators can rely on the assistant to remember past problems. On a team level, we might observe changes to social interactions, team cohesion, team performance, and team structure. It's crucial to measure these direct and indirect effects as introducing conversational AI could (indirectly) cause reduced operator well-being if they have fewer human-human interactions. On an organizational level, long-term studies could measure key performance indicators such as production performance. By measuring and observing the long-term impacts of using conversational AI for knowledge-sharing in factories, practitioners and policymakers can make more informed decisions regarding their deployment.

CHANGE MANAGEMENT STRATEGY

Assuming the result of the long-term impact studies are positive, and anticipating the difficulties in introducing a disruptive tool such as an LLM-powered conversational AI, we suggest formulating a change management strategy that breaks down the introduction into incremental steps. This would allow for the minimization of risk, enable more manageable shifts in work practices and culture, and, if necessary, make adjustments to the tool or change management strategy. This change management strategy is more likely to succeed than expecting everyone to drastically change their way of working from one to another no matter how large the potential benefit, especially with complex socio-technical systems [41]. Defining a change management process will involve further analysis of current work culture and practices such that hurdles can be identified and overcome systematically with (incremental) changes and appropiate behavioral change strategies.

6.4.2. BROADER FUTURE WORK

Future work beyond the scope of this dissertation explores broader, innovative research areas that extend the foundational insights provided by the thesis.

System Design Challenges for AI-Mediated Knowledge Sharing

Building on the system designs developed in this dissertation, many design and HCI challenges persist, warranting deeper exploration. For example: when and how should the assistant ask clarifying questions before answering a user query? When and how should users (be asked) to share their knowledge? And how should critical decision-making be encouraged? Determining what information is missing to provide a high-quality answer is a natural skill for humans but LLM-powered information retrieval mechanisms such as retrieval augmented generation do not clarify questions by default. Future work could explore mechanisms to identify if additional information is needed to provide a high-quality answer and if and how it would be appropriate to ask, building on work by Deng *et al.* [270]. Regarding the capture of knowledge, new interactions could be explored, such as providing operators wearables for short audio recordings or context-aware questioning during task execution as an apprentice might do. This direction is supported by recent work showing the potential for LLM-powered tools to revise and summarize dictated text for later use [271]. Lastly, for encouraging critical decision-making—which is relevant for many domains considering the increasing use of AI decision support—future research could explore how LLM responses can be formulated to support critical thinking, building on strategies such as reflection [241] and appropriate AI explanations [242]. Addressing these design challenges will support the development of human-centered LLM-facilitated KS for manufacturing and other domains.

Beyond AI-Facilitated Knowledge Sharing: Long-Term Implications and Challenges

Technological advances such as LLMs can be disruptive [272], and it is not easy to imagine how their widespread integration and adoption might shift work practices [273] or learning outcomes [274]. This dissertation explores only one way of using LLMs, yet countless alternatives exist, and as the technology progresses, more will emerge. We believe a deeper understanding of the long-term effects of using LLM-powered tools for KS is key to informing future work from low-level design decisions to legislation. As often happens when a transformative tool is introduced, certain human expertise becomes less important, allowing for the development of others, such as how widespread factory automation shifted manufacturing work from physical and repetitive tasks to multidisciplinary knowledgeable production line operations [275]. Additionally, certain tasks can be completed quicker, improving overall productivity—at least in theory. Based on understanding the effects of using LLM tools for knowledge-sharing nowadays, future work could employ speculative design to imagine what might be possible in 1, 2, or 5 years, such that we can make a more informed decision on where to focus our research efforts to ensure we are well prepared as a society to harness the technology for good and mitigate the eventual risks.

Framework to Reconcile Values Held by Operators, Managers, and Developers to Tackle Organizational and Cultural Change

We observed diverse goals and values across the stakeholders involved in our research on conversational AI in factories, similar to observations made by Baxter and Sommerville

[3]. While factory managers are concerned with how the system might add value to the organization, the operators are more concerned about making their work easier. The developers aim to deliver whatever system requirements are defined. We—as design engineers looking through a socio-technical lens—often found ourselves in a mediator-type position attempting to reconcile these different perspectives while also being motivated by our own research agenda and project requirements. We believe a deeper understanding of balancing these perspectives and implications would be valuable and support changes to the organizational structure and culture. For example, whether focusing on operator needs would eventually result in value added to the organization. This echoes prior work emphasizing the need for organizational structures and cultural changes to support digitized KS in manufacturing [6]. A framework to map these values and develop a strategy for reconciling them would be valuable when designing for (complex) socio-technical systems, such as AI-facilitated KS.

INCENTIVES FOR SHARING KNOWLEDGE THROUGH A COGNITIVE ASSISTANT

In this dissertation, we observed immense variability in the motivations of operators to share their knowledge with others. While some operators saw it as part of their job or enjoyed the process, others were indifferent or unmotivated to share. The topic of incentives for KS is widely discussed in the literature, especially in a work context [276–281], for example, by offering bonuses and to a lesser extent for crowdsourcing general knowledge, for example, by gamifying the experience [76]. However, sharing knowledge through conversational AI might impact the perceived benefits and efforts involved, requiring additional investigation. The incentives challenge is closely linked to the perceived effort of sharing knowledge and information and if this is rewarding in itself. Therefore, it makes sense to research these aspects holistically, for example, to investigate if reducing needed interaction time impacts user acceptance or the need for additional incentive mechanisms. Furthermore, there are gaps in the literature regarding incentives for sharing knowledge in other (non-work) communities, such as residents in low-resource neighborhoods sharing practical tips or lifestyle advice. Building on prior work that explored the role of incentives via digital mediums such as enterprise social-media [282], future work could explore the factors when facilitated by a CA in a factory context. Overall, incentives—whether monetary or social—represent a key component for any successful KS project, warranting focused research.

EXPLORING TRANSFERABILITY TO OTHER CONTEXTS AND DOMAINS

We suggest researching the transferability of conversational AI KS across different sizes of manufacturing operations and other contexts. Facilitating organizational or community memory is a widespread challenge across many industries and contexts, perhaps even universally, from hospitals to residential communities to space programs. Future work could explore how the systems, insights, and approaches presented in this dissertation can be transferred to other contexts. For example, to aid students during their master thesis project, a CA could support them by harnessing the community memory of prior and current students while also supporting the retrieval of official procedures. Investigating transferability stands to enhance the impact of this work, making it accessible to other domains.

6.5. CLOSING REMARKS

In this dissertation, we explored the integration of CAs in factory settings to enhance KS among operators. Our research identified both the potential benefits, challenges, and risks of deploying advanced AI technologies in such environments. We highlighted key design opportunities, user interaction strategies, privacy concerns, and the balance required between technological, social, organizational, and human factors for successful CA implementation.

Our findings show that CAs can improve efficiency, learning, and problem-solving on the production floor. However, they also emphasize the need for designs that prioritize human needs, organizational context, and ethical considerations that include the perspectives of all stakeholders. Future research should continue to explore the long-term impacts of CAs, refine their designs, and investigate the applicability of our findings in other settings.

In conclusion, while integrating CAs in factories offers a promising way to improve KS and operational efficiency, it is crucial to approach this integration with a thorough understanding of the socio-technical factors involved. Only by focusing on human-centered design and iterative field testing can we fully leverage the potential of CAs to create more efficient, knowledgeable, and adaptive manufacturing environments.

Bibliography

[1] C. Cimini, A. Lagorio, and P. Gaiardelli. “The Evolution of Operators’ Role in Production: How Lean Manufacturing and Industry 4.0 Affect Job Enlargement and Job Enrichment”. In: *International Journal of Production Research* 61.24 (Dec. 17, 2023), pp. 8493–8511. ISSN: 0020-7543. DOI: 10.1080/00207543.2022.2152894.

[2] D. Wurhofer, T. Meneweger, V. Fuchsberger, and M. Tscheligi. “Reflections on Operators’ and Maintenance Engineers’ Experiences of Smart Factories”. In: *Proceedings of the 2018 ACM International Conference on Supporting Group Work*. GROUP ’18. New York, NY, USA: Association for Computing Machinery, Jan. 7, 2018, pp. 284–296. ISBN: 978-1-4503-5562-9. DOI: 10.1145/3148330.3148349.

[3] G. Baxter and I. Sommerville. “Socio-technical systems: From design methods to systems engineering”. In: *Interacting with Computers* 23.1 (2011), pp. 4–17. ISSN: 0953-5438. DOI: https://doi.org/10.1016/j.intcom.2010.07.003.

[4] T.-T. (Huang, L. Chen, and R. A. Stewart. “The Moderating Effect of Knowledge Sharing on the Relationship between Manufacturing Activities and Business Performance”. In: *Knowledge Management Research & Practice* 8.4 (Dec. 1, 2010), pp. 285–306. ISSN: 1477-8246. DOI: 10.1057/kmrp.2010.21.

[5] S. Hoffmann, T. Ludwig, F. Jasche, V. Wulf, and D. Randall. “RetrofittAR: Supporting Hardware-Centered Expertise Sharing in Manufacturing Settings through Augmented Reality”. In: *Comput. Supported Coop. Work* 32.1 (June 2022), pp. 93–139. ISSN: 0925-9724. DOI: 10.1007/s10606-022-09430-x.

[6] D. Li, Å. Fast-Berglund, and D. Paulin. “Current and future Industry 4.0 capabilities for information and knowledge sharing: Case of two Swedish SMEs”. In: *The International Journal of Advanced Manufacturing Technology* 105.9 (2019), pp. 3951–3963. DOI: 10.1007/s00170-019-03942-5.

[7] I. Becerra-Fernandez and R. Sabherwal. *Knowledge management: Systems and processes*. Routledge, 2014.

[8] I. Nonaka and H. Takeuchi. “The knowledge-creating company: How Japanese companies create the dynamics of innovation”. In: *New York, NY* (1995).

[9] *Labour Force and Skills Shortages: How to Tackle Them?* Policy Orientation Note. Business Europe, 2022.

[10] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, W. Ye, Y. Zhang, Y. Chang, P. S. Yu, Q. Yang, and X. Xie. “A Survey on Evaluation of Large Language Models”. In: *ACM Transactions on Intelligent Systems and Technology* 15.3 (Mar. 29, 2024), 39:1–39:45. ISSN: 2157-6904. DOI: 10.1145/3641289.

[11] *Commission Actions Tackle Labour and Skills Shortages in EU*. European Commission. Apr. 20, 2024.

[12] D. A. Waterman. *A guide to expert systems*. Addison-Wesley Longman Publishing Co., Inc., 1985.

[13] C. Wagner. "Breaking the Knowledge Acquisition Bottleneck Through Conversational Knowledge Management". In: *Information Resources Management Journal (IRMJ)* 19.1 (Jan. 1, 2006), pp. 70–83. ISSN: 1040-1628. DOI: 10.4018/irmj.2006010104.

[14] H. C. Tan, P. Carrillo, C. Anumba, J. M. Kamara, D. Bouchlaghem, and C. Udeaja. "Live Capture and Reuse of Project Knowledge in Construction Organisations". In: *Knowledge Management Research & Practice* 4.2 (May 1, 2006), pp. 149–161. ISSN: 1477-8238. DOI: 10.1057/palgrave.kmrp.8500097.

[15] M. Müller, E. Alexandi, and J. Metternich. "Digital Shop Floor Management Enhanced by Natural Language Processing". In: *Procedia CIRP*. 8th CIRP Global Web Conference – Flexible Mass Customisation (CIRPe 2020) 96 (Jan. 1, 2021), pp. 21–26. ISSN: 2212-8271. DOI: 10.1016/j.procir.2021.01.046.

[16] B. Edwards, M. Zatorsky, and R. Nayak. "Clustering and classification of maintenance logs using text data mining". In: *Volume 87-Data Mining and Analytics 2008* (2008), pp. 193–199.

6

[17] O. Oruç. "A semantic Question Answering through Heterogeneous data source in the domain of smart factory". In: *International Journal on Natural Language Computing (IJNLC) Vol* 9 (2020).

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc., 2017.

[19] M. Casillo, F. Colace, L. Fabbri, M. Lombardi, A. Romano, and D. Santaniello. "Chatbot in Industry 4.0: An Approach for Training New Employees". In: *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. 2020, pp. 371–376. DOI: 10.1109/TALE48869.2020.9368339.

[20] S. Mantravadi, A. D. Jansson, and C. Moller. "User-Friendly MES Interfaces: Recommendations for an AI-Based Chatbot Assistance in Industry 4.0 Shop Floors". In: *Intelligent Information and Database Systems (ACIIDS 2020), PT II*. 12th Asian Conference on Intelligent Information and Database Systems (ACIIDS). Ed. by N. T. Nguyen, K. Jearanaitanakij, A. Selamat, B. Trawinski, and S. Chittayasothorn. Vol. 12034. Cham: Springer International Publishing Ag, 2020, pp. 189–201. DOI: 10.1007/978-3-030-42058-1_16.

[21] A. J. C. Trappey, C. V. Trappey, M.-H. Chao, and C.-T. Wu. "VR-enabled Engineering Consultation Chatbot for Integrated and Intelligent Manufacturing Services". In: *Journal of Industrial Information Integration* 26 (Mar. 1, 2022), p. 100331. ISSN: 2452-414X. DOI: 10.1016/j.jii.2022.100331.

[22] H. Chen, F. Jiao, X. Li, C. Qin, M. Ravaut, R. Zhao, C. Xiong, and S. Joty. *ChatGPT's One-year Anniversary: Are Open-Source Large Language Models Catching up?* 2024. DOI: 10.48550/arXiv.2311.16989. arXiv: 2311.16989 [cs.CL]. preprint.

[23] S. Wellsandt, Z. Rusak, S. Ruiz Arenas, D. Aschenbrenner, K. A. Hribernik, and K.-D. Thoben. *Concept of a Voice-Enabled Digital Assistant for Predictive Maintenance in Manufacturing*. Oct. 23, 2020. DOI: 10.2139/ssrn.3718008.

[24] N. Gkatzios, N. Vryonis, C. Fragkos, C. Sakkas, V. Mavrikakis, V. Koumaras, G. Makropoulos, D. Fragkos, and H. Koumaras. "A Chatbot Assistant for Optimizing the Fault Detection and Diagnostics of Industry 4.0 Equipment in the 6G Era". In: *2023 IEEE Conference on Standards for Communications and Networking, CSCN*. IEEE Conference on Standards for Communications and Networking (CSCN). New York: IEEE, 2023, pp. 131–136. DOI: 10.1109/CSCN60443.2023.10453129.

[25] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin. "Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges". In: *IEEE Access* 6 (2018), pp. 6505–6519. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2783682.

[26] E. Kaasinen, F. Schmalfuß, C. Özturk, S. Aromaa, M. Boubekeur, J. Heilala, P. Heikkilä, T. Kuula, M. Liinasuo, S. Mach, R. Mehta, E. Petäjä, and T. Walter. "Empowering and Engaging Industrial Workers with Operator 4.0 Solutions". In: *Computers & Industrial Engineering* 139 (Jan. 1, 2020), p. 105678. ISSN: 0360-8352. DOI: 10.1016/j.cie.2019.01.052.

[27] L. Hannola, F. Lacueva-Pérez, P. Pretto, A. Richter, M. Schafler, and M. Steinhüser. "Assessing the Impact of Socio-Technical Interventions on Shop Floor Work Practices". In: *International Journal of Computer Integrated Manufacturing* 33.6 (June 2, 2020), pp. 550–571. ISSN: 0951-192X. DOI: 10.1080/0951192X.2020.1775296.

[28] S. Nahavandi. "Industry 5.0—A human-centric solution". In: *Sustainability* 11.16 (2019), p. 4371. DOI: 10.3390/su11164371.

[29] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang. "Industry 4.0 and Industry 5.0—Inception, conception and perception". In: *Journal of Manufacturing Systems* 61 (2021), pp. 530–535. ISSN: 02786125. DOI: 10.1016/j.jmsy.2021.10.006.

[30] A. Sheth, H. Y. Yip, A. Iyengar, and P. Tepper. "Cognitive services and intelligent chatbots: current perspectives and special issue introduction". In: *IEEE Internet Computing* 23.2 (2019), pp. 6–12. DOI: 10.1109/MIC.2018.2889231.

[31] S. Hoffmann, A. F. P. de Carvalho, D. Abele, M. Schweitzer, P. Tolmie, and V. Wulf. "Cyber-Physical Systems for Knowledge and Expertise Sharing in Manufacturing Contexts: Towards a Model Enabling Design". In: *Computer Supported Cooperative Work* 28.3-4 (June 1, 2019), pp. 469–509. ISSN: 0925-9724. DOI: 10.1007/s10606-019-09355-y.

[32] S. Durst and S. Wilhelm. "Knowledge Management in Practice: Insights into a Medium-sized Enterprise's Exposure to Knowledge Loss". In: *Prometheus* 29.1 (Mar. 1, 2011). ISSN: 0810-9028, 1470-1030. DOI: 10.1080/08109028.2011.565693.

[33] P. Wellener, V. Reyes, H. Ashton, and C. Moutray. *Creating Pathways for Tomorrow's Workforce Today: Beyond Reskilling in Manufacturing*. Deloitte, The Manufacturing Institute, 2021.

[34] E. C. Martins and N. Martins. "The Role of Organisational Factors in Combating Tacit Knowledge Loss in Organisations". In: *Southern African Business Review* 15.1 (1 2011). ISSN: 1998-8125.

[35] M. ł. Zięba. "Knowledge Safety–Insights from the SME Sector". In: *Central European Management Journal* 25.3 (2017), pp. 78–96.

[36] Y. Cheng, Y. Zhang, P. Ji, W. Xu, Z. Zhou, and F. Tao. "Cyber-Physical Integration for Moving Digital Factories Forward towards Smart Manufacturing: A Survey". In: *The International Journal of Advanced Manufacturing Technology* 97.1 (July 1, 2018), pp. 1209–1221. DOI: 10.1007/s00170-018-2001-2.

[37] F. Mörike. "Inverted Hierarchies on the Shop Floor: The Organisational Layer of Workarounds for Collaboration in the Metal Industry". In: *Computer Supported Cooperative Work* 31.1 (Mar. 1, 2022), pp. 111–147. ISSN: 0925-9724. DOI: 10.1007/s10606-021-09415-2.

[38] L. Hannola, A. Richter, S. Richter, and A. Stocker. "Empowering Production Workers with Digitally Facilitated Knowledge Processes – a Conceptual Framework". In: *International Journal of Production Research* 56.14 (July 18, 2018), pp. 4729–4743. ISSN: 0020-7543. DOI: 10.1080/00207543.2018.1445877.

6

[39] F. Longo, L. Nicoletti, and A. Padovano. "Ubiquitous Knowledge Empowers the Smart Factory: The Impacts of a Service-oriented Digital Twin on Enterprises' Performance". In: *Annual Reviews in Control* 47 (Jan. 1, 2019), pp. 221–236. ISSN: 1367-5788. DOI: 10.1016/j.arcontrol.2019.01.001.

[40] A. M. Abubakar, H. Elrehail, M. A. Alatailat, and A. Elçi. "Knowledge Management, Decision-Making Style and Organizational Performance". In: *Journal of Innovation & Knowledge* 4.2 (Apr. 1, 2019), pp. 104–114. ISSN: 2444-569X. DOI: 10.1016/j.jik.2017.07.003.

[41] D. A. Norman and P. J. Stappers. "DesignX: Complex Sociotechnical Systems". In: *She Ji: The Journal of Design, Economics, and Innovation* 1.2 (Dec. 1, 2015), pp. 83–106. ISSN: 2405-8726. DOI: 10.1016/j.sheji.2016.01.002.

[42] T. Sexton, M. P. Brundage, M. Hoffman, and K. C. Morris. "Hybrid datafication of maintenance logs from AI-assisted human tags". In: *2017 IEEE International Conference on Big Data (Big Data)*. 2017, pp. 1769–1777. DOI: 10.1109/BigData.2017.8258120.

[43] N. J. Cooke. "Varieties of Knowledge Elicitation Techniques". In: *International Journal of Human-Computer Studies* 41.6 (Dec. 1, 1994), pp. 801–849. ISSN: 1071-5819. DOI: 10.1006/ijhc.1994.1083.

[44] I. Nonaka. "The knowledge-creating company". In: *The economic impact of knowledge*. Routledge, 2009, pp. 175–187.

[45] M. Alavi and D. E. Leidner. "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues". In: *MIS Quarterly* 25.1 (2001), pp. 107–136. ISSN: 0276-7783. DOI: 10.2307/3250961. JSTOR: 3250961.

[46] G. J. B. Probst, S. Raub, and K. Romhardt. *Managing Knowledge: Building Blocks for Success*. Wiley, Jan. 4, 2000. 368 pp. ISBN: 978-0-471-99768-9. Google Books: aXyFQgAACAAJ.

[47] A. P. Massey, M. M. Montoya-Weiss, and K. Holcom. "Re-Engineering the Customer Relationship: Leveraging Knowledge Assets at IBM". In: *Decision Support Systems*. Decision Support Issues in Customer Relationship Management and Interactive Marketing for E-Commerce 32.2 (Dec. 1, 2001), pp. 155–170. DOI: 10.1016/S0167-9236(01)00108-7.

[48] K. R. Lakhani, K. Hutter, S. H. Pokrywa, and J. Fuller. *Open Innovation at Siemens*. Case Study 613-100. Harvard Business School, 2013.

[49] J. Liebowitz. "A look at NASA Goddard Space Flight Center's knowledge management initiatives". In: *IEEE Software* 19.3 (2002), pp. 40–42. DOI: 10.1109/MS.2002.1003451.

[50] C. Duhigg. "What Google learned from its quest to build the perfect team". In: *The New York Times Magazine* 26.2016 (2016), p. 2016.

[51] J. H. Dyer and K. Nobeoka. "Creating and managing a high-performance knowledge-sharing network: the Toyota case". en. In: *Strategic Management Journal* 21.3 (2000), pp. 345–367. ISSN: 1097-0266. DOI: 10.1002/(SICI)1097-0266(200003)21:3<345::AID-SMJ96>3.0.CO;2-N.

[52] J. S. Brown and P. Duguid. "Balancing act: How to capture knowledge without killing it". In: *Harvard business review* 78.3 (2000), pp. 73–80.

[53] D. G. Bobrow and J. Whalen. "Community knowledge sharing in practice: the Eureka story". In: *Reflections: The SoL Journal* 4.2 (2002), pp. 47–59.

[54] J. Cullen and A. Bryman. "The knowledge acquisition bottleneck: time for reassessment?" In: *Expert Systems* 5.3 (1988), pp. 216–225. DOI: 10.1111/j.1468-0394.1988.tb00065.x.

[55] E. Strickland. "IBM Watson, Heal Thyself: How IBM Overpromised and Underdelivered on AI Health Care". In: *IEEE Spectrum* 56.4 (Apr. 2019), pp. 24–31. DOI: 10.1109/MSPEC.2019.8678513.

[56] E. Ras, F. Wild, C. Stahl, and A. Baudet. "Bridging the Skills Gap of Workers in Industry 4.0 by Human Performance Augmentation Tools: Challenges and Roadmap". In: *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*. PETRA '17. New York, NY, USA: Association for Computing Machinery, June 21, 2017, pp. 428–432. ISBN: 978-1-4503-5227-7. DOI: 10.1145/3056540.3076192.

[57] S. Colabianchi, A. Tedeschi, and F. Costantino. "Human-Technology Integration with Industrial Conversational Agents: A Conceptual Architecture and a Taxonomy for Manufacturing". In: *Journal of Industrial Information Integration* 35 (Oct. 1, 2023), p. 100510. ISSN: 2452-414X. DOI: 10.1016/j.jii.2023.100510.

[58] L. Hoerner, M. Schamberger, and F. Bodendorf. "Using Tacit Expert Knowledge to Support Shop-floor Operators Through a Knowledge-based Assistance System". en. In: *Computer Supported Cooperative Work (CSCW)* (Sept. 2022). ISSN: 1573-7551. DOI: 10.1007/s10606-022-09445-4.

[59] K. S. Kiangala and Z. Wang. "An Experimental Hybrid Customized AI and Generative AI Chatbot Human Machine Interface to Improve a Factory Troubleshooting Downtime in the Context of Industry 5.0". In: *The International Journal of Advanced Manufacturing Technology* (Apr. 3, 2024). ISSN: 0268-3768, 1433-3015. DOI: 10.1007/s00170-024-13492-0.

[60] G. Castane, A. Dolgui, N. Kousi, B. Meyers, S. Thevenin, E. Vyhmeister, and P.-O. Ostberg. "The ASSISTANT Project: AI for High Level Decisions in Manufacturing". In: *International Journal of Production Research* 61.7 (Apr. 3, 2023), pp. 2288–2306. ISSN: 0020-7543, 1366-588X. DOI: 10.1080/00207543.2022.2069525.

[61] W. Tao, Z.-H. Lai, M. C. Leu, Z. Yin, and R. Qin. "A self-aware and active-guiding training & assistant system for worker-centered intelligent manufacturing". en. In: *Manufacturing Letters* 21 (Aug. 2019), pp. 45–49. ISSN: 2213-8463. DOI: 10.1016/j.mfglet.2019.08.003.

[62] A. Maedche, C. Legner, A. Benlian, B. Berger, H. Gimpel, T. Hess, O. Hinz, S. Morana, and M. Söllner. "AI-based digital assistants: Opportunities, threats, and research perspectives". In: *Business & Information Systems Engineering* 61 (2019), pp. 535–544. DOI: 10.1007/s12599-019-00600-8.

[63] M. A. Dhuieb, F. Laroche, and A. Bernard. "Digital Factory Assistant: Conceptual Framework and Research Propositions". In: *Product Lifecycle Management for Society (PLM 2013)*. 10th IFIP WG 5.1 International Conference on Product Lifecycle Management (PLM). Ed. by A. Bernard, L. Rivest, and D. Dutta. Vol. 409. Berlin: Springer-Verlag Berlin, 2013, pp. 500–509. ISBN: 978-3-642-41501-2.

[64] P. Berger and J. von Garrel. "How to Design a Value-Based Chatbot for the Manufacturing Industry: An Empirical Study of an Internal Assistance for Employees". In: *Kunstliche Intelligenz* (Nov. 13, 2023). ISSN: 0933-1875, 1610-1987. DOI: 10.1007/s13218-023-00817-6.

[65] C. Li, A. K. Hansen, D. Chrysostomou, S. Bogh, and O. Madsen. "Bringing a Natural Language-enabled Virtual Assistant to Industrial Mobile Robots for Learning, Training and Assistance of Manufacturing Tasks". In: *2022 IEEE/SICE International Symposium on System Integration (SII 2022)*. IEEE/SICE International Symposium on System Integration (SII). New York: IEEE, 2022, pp. 238–243. ISBN: 978-1-66544-540-5. DOI: 10.1109/SII52469.2022.9708757.

[66] S. Wellsandt, K. Klein, K. Hribernik, M. Lewandowski, A. Bousdekis, G. Mentzas, and K.-D. Thoben. "Hybrid-Augmented Intelligence in Predictive Maintenance with Digital Intelligent Assistants". In: *Annual Reviews in Control* 53 (Jan. 1, 2022), pp. 382–390. ISSN: 1367-5788. DOI: 10.1016/j.arcontrol.2022.04.001.

[67] S. M. R. Naqvi, M. Ghufran, S. Meraghni, C. Varnier, J.-M. Nicod, and N. Zerhouni. "Human Knowledge Centered Maintenance Decision Support in Digital Twin Environment". In: *Journal of Manufacturing Systems* 65 (Oct. 2022), pp. 528–537. ISSN: 0278-6125, 1878-6642. DOI: 10.1016/j.jmsy.2022.10.003.

[68] D. Rooein, D. Bianchini, F. Leotta, M. Mecella, P. Paolini, and B. Pernici. "Chatting About Processes in Digital Factories: A Model-Based Approach". In: *Enterprise, Business-Process and Information Systems Modeling*. Ed. by S. Nurcan, I. Reinhartz-Berger, P. Soffer, and J. Zdravkovic. Cham: Springer International Publishing, 2020, pp. 70–84. ISBN: 978-3-030-49418-6. DOI: 10.1007/978-3-030-49418-6_5.

[69] M. Wieland, P. Hirmer, F. Steimle, C. Groeger, B. Mitschang, E. Rehder, D. Lucke, O. A. Rahman, and T. Bauernhansl. "Towards a Rule-Based Manufacturing Integration Assistant". In: *Factories of the Future in the Digital Environment*. 49th CIRP Conference on Manufacturing Systems (CIRP-CMS). Ed. by E. Westkamper and T. Bauernhansl. Vol. 57. Amsterdam: Elsevier Science Bv, 2016, pp. 213–218. DOI: 10.1016/j.procir.2016.11.037.

[70] C. Li and H. J. Yang. "Bot-X: An AI-based Virtual Assistant for Intelligent Manufacturing". In: *Multiagent and Grid Systems* 17.1 (2021), pp. 1–14. ISSN: 1574-1702, 1875-9076. DOI: 10.3233/MGS-210340.

[71] A. Reis, J. Barroso, A. Santos, P. Rodrigues, and R. Pereira. "Virtual Assistance in the Context of the Industry 4.0: A Case Study at Continental Advanced Antenna". In: *Information Systems and Technologies*. Ed. by A. Rocha, H. Adeli, G. Dzemyda, and F. Moreira. Cham: Springer International Publishing, 2022, pp. 651–662. ISBN: 978-3-031-04826-5. DOI: 10.1007/978-3-031-04826-5_64.

[72] D. Pavlov, I. Sosnovsky, V. Dimitrov, V. Melentyev, and D. Korzun. "Case Study of Using Virtual and Augmented Reality in Industrial System Monitoring". In: *2020 26th Conference of Open Innovations Association (FRUCT)*. 2020 26th Conference of Open Innovations Association (FRUCT). Apr. 2020, pp. 367–375. DOI: 10.23919/FRUCT48808.2020.9087410.

[73] M. Baldauf, S. Müller, A. Seeliger, T. Küng, A. Michel, and W. Züllig. "Human Interventions in the Smart Factory – A Case Study on Co-Designing Mobile and Wearable Monitoring Systems with Manufacturing Staff". In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI EA '21. New York, NY, USA: Association for Computing Machinery, May 8, 2021, pp. 1–6. ISBN: 978-1-4503-8095-9. DOI: 10.1145/3411763.3451774.

[74] C. Angulo, A. Chacón, and P. Ponsa. "Towards a Cognitive Assistant Supporting Human Operators in the Artificial Intelligence of Things". In: *Internet of Things* 21 (Apr. 1, 2023), p. 100673. ISSN: 2542-6605. DOI: 10.1016/j.iot.2022.100673.

[75] E. Fenoglio, E. Kazim, H. Latapie, and A. Koshiyama. "Tacit knowledge elicitation process for industry 4.0". en. In: *Discover Artificial Intelligence* 2.1 (Mar. 2022), p. 6. ISSN: 2731-0809. DOI: 10.1007/s44163-022-00020-w.

[76] A. Balayn, G. He, A. Hu, J. Yang, and U. Gadiraju. "Ready Player One! Eliciting Diverse Knowledge Using A Configurable Game". In: *Proceedings of the ACM Web Conference 2022*. WWW '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 1709–1719. ISBN: 9781450390965. DOI: 10.1145/3485447.3512241.

[77] L. Bradeško, M. Witbrock, J. Starc, Z. Herga, M. Grobelnik, and D. Mladenić. "Curious Cat–Mobile, Context-Aware Conversational Crowdsourcing Knowledge Acquisition". In: *ACM Trans. Inf. Syst.* 35.4 (Aug. 2017). ISSN: 1046-8188. DOI: 10.1145/3086686.

[78] C. Kotthaus, N. Vitt, M. Krüger, V. Pipek, and V. Wulf. "Negotiating Priorities on the Shopfloor: A Design Case Study of Maintainers' Practices". In: *Comput. Supported Coop. Work* 32.1 (Oct. 2022), pp. 141–210. ISSN: 0925-9724. DOI: 10.1007/s10606-022-09444-5.

[79] T. Ludwig, A. Boden, and V. Pipek. "3D Printers as Sociable Technologies: Taking Appropriation Infrastructures to the Internet of Things". In: 24.2 (Apr. 2017). ISSN: 1073-0516. DOI: 10.1145/3007205.

6

[80] F. Cabitza, A. Campagner, and C. Simone. "The need to move away from agential-AI: Empirical investigations, useful concepts and open issues". In: *International Journal of human-computer studies* 155 (2021), p. 102696. DOI: 10.1016/j.ijhcs.2021.102696.

[81] P. Madhavan, D. A. Wiegmann, and F. C. Lacson. "Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids". In: *Human Factors* 48.2 (June 1, 2006), pp. 241–256. ISSN: 0018-7208. DOI: 10.1518/001872006777724408.

[82] M. Ezzamel, H. Willmott, and F. Worthington. "Power, Control and Resistance in 'The Factory That Time Forgot'". In: *Journal of Management Studies* 38.8 (2001), pp. 1053–1079. ISSN: 1467-6486. DOI: 10.1111/1467-6486.00272.

[83] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. V. Ess-Dykema, and M. Meteer. "Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech". In: *Computational Linguistics* 26.3 (Sept. 2000), pp. 339–373. ISSN: 0891-2017. DOI: 10.1162/089120100561737. eprint: https://direct.mit.edu/coli/article-pdf/26/3/339/1797531/089120100561737.pdf.

[84] B. Luo, R. Y. Lau, C. Li, and Y.-W. Si. "A critical review of state-of-the-art chatbot designs and applications". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 12.1 (2022), e1434. DOI: 10.1002/widm.1434.

[85] A. M. Rahman, A. A. Mamun, and A. Islam. "Programming challenges of chatbot: Current and future prospective". In: *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. 2017, pp. 75–78. DOI: 10.1109/R10-HTC.2017.8288910.

[86] S. Karita, N. Chen, T. Hayashi, T. Hori, H. Inaguma, Z. Jiang, M. Someki, N. E. Y. Soplin, R. Yamamoto, X. Wang, S. Watanabe, T. Yoshimura, and W. Zhang. "A Comparative Study on Transformer vs RNN in Speech Applications". In: *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. 2019, pp. 449–456. DOI: 10.1109/ASRU46091.2019.9003750.

[87] G. Jawahar, B. Sagot, and D. Seddah. "What Does BERT Learn about the Structure of Language?" In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 3651–3657. DOI: 10.18653/v1/P19-1356.

[88] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.-Y. Nie, and J.-R. Wen. *A Survey of Large Language Models*. 2023. arXiv: 2303.18223 [cs.CL]. preprint.

[89] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus. *Emergent Abilities of Large Language Models*. 2022. arXiv: 2206.07682 [cs.CL]. preprint.

[90] N. Karanikolas, E. Manga, N. Samaridi, E. Tousidou, and M. Vassilakopoulos. "Large Language Models versus Natural Language Understanding and Generation". In: *Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics*. PCI '23. New York, NY, USA: Association for Computing Machinery, Feb. 14, 2024, pp. 278–290. DOI: 10.1145/3635059.3635104.

[91] F. Leiser, S. Eckhardt, V. Leuthe, M. Knaeble, A. Mädche, G. Schwabe, and A. Sunyaev. "HILL: A Hallucination Identifier for Large Language Models". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. Honolulu, HI, USA: Association for Computing Machinery, 2024. DOI: 10.1145/3613904.3642428.

[92] Mercedes-Benz. *Benz tests chatgpt in intelligent vehicle production.: Mercedes-Benz Group*. July 2023.

[93] Y. Xia, M. Shenoy, N. Jazdi, and M. Weyrich. *Towards autonomous system: flexible modular production system enhanced with large language model agents*. 2023. arXiv: 2304.14721 [cs.RO]. preprint.

[94] X. Wang, N. Anwer, Y. Dai, and A. Liu. "ChatGPT for design, manufacturing, and education". In: *Procedia CIRP* 119 (2023). The 33rd CIRP Design Conference, pp. 7–14. ISSN: 2212-8271. DOI: https://doi.org/10.1016/j.procir.2023.04.001.

[95] Y. Bang, S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, H. Lovenia, Z. Ji, T. Yu, W. Chung, Q. V. Do, Y. Xu, and P. Fung. *A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity*. 2023. arXiv: 2302.04023 [cs.CL]. preprint.

6

[96] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. V. Le, and D. Zhou. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models". In: *Advances in Neural Information Processing Systems* 35 (Dec. 6, 2022), pp. 24824–24837.

[97] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. "Language Models are Few-Shot Learners". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901.

[98] T. Gao, A. Fisch, and D. Chen. "Making Pre-trained Language Models Better Few-shot Learners". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 3816–3830. DOI: 10.18653/v1/2021.acl-long.295.

[99] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks". In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS'20. Vancouver, BC, Canada: Curran Associates Inc., 2020. ISBN: 9781713829546.

[100] K. Schmidt and L. Bannon. "Taking CSCW seriously: Supporting articulation work". In: *Computer supported cooperative work (CSCW)* 1 (1992), pp. 7–40. DOI: 10.1007/bf00752449.

[101] B. Shneiderman. *Human-centered AI*. Oxford University Press, 2022.

[102] S. Kernan Freire, S. S. Panicker, S. Ruiz-Arenas, Z. Rusák, and E. Niforatos. "A Cognitive Assistant for Operators: AI-Powered Knowledge Sharing on Complex Systems". In: *IEEE Pervasive Computing* (2022), pp. 1–9. DOI: 10.1109/MPRV.2022.3218600.

[103] S. Kernan Freire, E. Niforatos, C. Wang, S. Ruiz-Arenas, M. Foosherian, S. Wellsandt, and A. Bozzon. "Lessons Learned from Designing and Evaluating CLAICA: A Continuously Learning AI Cognitive Assistant". In: *Proceedings of the 28th International Conference on Intelligent User Interfaces*. IUI '23. Sydney, NSW, Australia: Association for Computing Machinery, 2023, pp. 553–568. DOI: 10.1145/3581641.3584042.

[104] S. Benjamin, M. Marathamuthu, and U. Murugaiah. "The use of 5-WHYs technique to eliminate OEE's speed loss in a manufacturing firm". en. In: *J. Qual. Maint. Eng* 21.4 (Jan. 2015). Journal, pp. 419–435.

[105] S. Kernan Freire, C. Wang, M. Foosherian, S. Wellsandt, S. Ruiz-Arenas, and E. Niforatos. "Knowledge sharing in manufacturing using LLM-powered tools: user study and model benchmarking". In: *Frontiers in Artificial Intelligence* 7 (2024). ISSN: 2624-8212. DOI: 10.3389/frai.2024.1293084.

[106] J. Fereday and E. Muir-Cochrane. "Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development". In: *International journal of qualitative methods* 5.1 (2006), pp. 80–92. DOI: 10.1177/160940690600500107.

[107] J. Thomas and A. Harden. "Methods for the thematic synthesis of qualitative research in systematic reviews". In: *BMC medical research methodology* 8.1 (2008), pp. 1–10. DOI: 10.1186/1471-2288-8-45.

[108] S. Kernan Freire, T. He, C. Wang, E. Niforatos, and A. Bozzon. "Operators' Perspectives on Conversational AI for Knowledge Sharing: Challenges, Risks and Impact on Work". In: *Proceedings of the ACM on Human-Computer Interaction* (CSCW 2025). submitted.

[109] J. Howells. "Tacit Knowledge". In: *Technology Analysis & Strategic Management* 8.2 (Jan. 1, 1996), pp. 91–106. ISSN: 0953-7325. DOI: 10.1080/09537329608524237.

[110] I. Nonaka and H. Takeuchi. *The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation*. Vol. 105. Oxford University Press, 1995. 320 pp. ISBN: 978-0-19-509269-1.

[111] H. Mohajan. "Sharing of Tacit Knowledge in Organizations: A Review". In: *American Journal of Computer Science and Engineering* 3.2 (June 5, 2016), pp. 6–19.

[112] X. C. Chen, A. Sagar, J. T. Kao, T. Y. Li, C. Klein, S. Pulman, A. Garg, and J. D. Williams. "Active Learning for Domain Classification in a Commercial Spoken Personal Assistant". In: *Proc. Interspeech 2019*. 2019, pp. 1478–1482. DOI: 10.21437/Interspeech.2019-1315.

[113] S. K. Kondreddi, P. Triantafillou, and G. Weikum. "Combining information extraction and human computing for crowdsourced knowledge acquisition". In: *2014 IEEE 30th International Conference on Data Engineering*. 2014, pp. 988–999. DOI: 10.1109/ICDE.2014.6816717.

[114] J. Huang, M. Zhou, and D. Yang. "Extracting Chatbot Knowledge from Online Discussion Forums". In: *Proceedings of the 20th International Joint Conference on Artifical Intelligence*. IJCAI'07. Hyderabad, India: Morgan Kaufmann Publishers Inc., 2007, pp. 423–428.

[115] T. H. Davenport, P. D. o. M. I. S. o. M. T. H. Davenport, and L. Prusak. *Working Knowledge: How Organizations Manage What They Know*. Harvard Business Press, 1998. 224 pp. ISBN: 978-0-87584-655-2.

[116] C. Wagner. "Breaking the Knowledge Acquisition Bottleneck Through Conversational Knowledge Management". In: *Information Resource Management Journal* 19.1 (Jan. 2006), pp. 70–83. DOI: 10.4018/irmj.2006010104.

[117] S. Kernan Freire, E. Niforatos, Z. Rusak, D. Aschenbrenner, and A. Bozzon. "A Conversational User Interface for Instructional Maintenance Reports". In: *Proceedings of the 4th Conference on Conversational User Interfaces*. CUI '22. Glasgow, United Kingdom: Association for Computing Machinery, 2022. DOI: 10.1145/3543829.3544516.

[118] T. L. Johnson, S. R. Fletcher, W. Baker, and R. L. Charles. "How and Why We Need to Capture Tacit Knowledge in Manufacturing: Case Studies of Visual Inspection". In: *Applied Ergonomics* 74 (Jan. 1, 2019), pp. 1–9. ISSN: 0003-6870. DOI: 10.1016/j.apergo.2018.07.016.

[119] I. Nonaka and G. Von Krogh. "Perspective—Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory". en. In: *Organization science* 20.3 (Jan. 2009), pp. 635–652. DOI: 10.1287/orsc.1080.0412.

[120] H. Hashimoto, I. Yoshida, Y. Teramoto, H. Tabata, and C. Han. "Extraction of Tacit Knowledge as Expert Engineer's Skill Based on Mixed Human Sensing". In: *2011 RO-MAN*. 2011 RO-MAN. July 2011, pp. 413–418. DOI: 10.1109/ROMAN.2011.6005234.

[121] J. Schmid, S. Trabesinger, M. Brillinger, R. Pichler, J. Wurzinger, and R. Ciumasu. "Tacit Knowledge Based Acquisition of Verified Machining Data". In: *2020 9th International Conference on Industrial Technology and Management (ICITM)*. 2020 9th International Conference on Industrial Technology and Management (ICITM). Feb. 2020, pp. 117–121. DOI: 10.1109/ICITM48982.2020.9080346.

[122] Q. Wan, H. Zhang, and Y. Wang. "Tacit Knowledge Representation and Classification Based on Statistical Data for Production Process". In: *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). May 2021, pp. 655–660. DOI: 10.1109/CSCWD49262.2021.9437840.

[123] Å. Fast-Berglund, P. Thorvald, E. Billing, A. Palmquist, D. Romero, and G. Weichhart. "Conceptualizing Embodied Automation to Increase Transfer of Tacit Knowledge in the Learning Factory". In: *2018 International Conference on Intelligent Systems (IS)*. 2018 International Conference on Intelligent Systems (IS). Sept. 2018, pp. 358–364. DOI: 10.1109/IS.2018.8710482.

[124] S. Katalnikova and L. Novickis. "Choice of Knowledge Representation Model for Development of Knowledge Base: Possible Solutions". In: *International Journal of Advanced Computer Science and Applications (IJACSA)* 9.2 (2 2018). ISSN: 2156-5570. DOI: 10.14569/IJACSA.2018.090249.

[125] T. de Jong and M. G. M. Ferguson-Hessler. "Cognitive Structures of Good and Poor Novice Problem Solvers in Physics". In: *Journal of educational psychology* 78.4 (1986), pp. 279–288. ISSN: 0022-0663. DOI: 10.1037/0022-0663.78.4.279.

[126] S. J. Benjamin, M. S. Marathamuthu, and U. Murugaiah. "The Use of 5-WHYs Technique to Eliminate OEE's Speed Loss in a Manufacturing Firm". In: *Journal of Quality in Maintenance Engineering* 21.4 (Jan. 1, 2015), pp. 419–435. ISSN: 1355-2511. DOI: 10.1108/JQME-09-2013-0062.

[127] S. Durst and M. Zieba. "Mapping Knowledge Risks: Towards a Better Understanding of Knowledge Management". In: *Knowledge Management Research & Practice* 17.1 (Jan. 2, 2019), pp. 1–13. ISSN: 1477-8238. DOI: 10.1080/14778238.2018.1538603.

[128] K.-D. Thoben, S. Wiesner, and T. Wuest. ""Industrie 4.0" and Smart Manufacturing – A Review of Research Issues and Application Examples". In: *International Journal of Automation Technology* 11.1 (2017), pp. 4–16. ISSN: 1881-7629. DOI: 10.20965/ijat.2017.p0004.

[129] J. Müller. "Enabling Technologies for Industry 5.0—Results of a Workshop with Europe's Technology Leaders". In: *Directorate-General for Research and Innovation* (2020). DOI: 10.2777/082634.

[130] D. Romero, J. Stahre, T. Wuest, O. Noran, P. Bernus, Å. Fast-Berglund, and D. Gorecky. "Towards an operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies". In: *Proceedings of the 46th International Conference on Computers and Industrial Engineering*. Ed. by M. Dessouky, Y. Dessouky, and H. K. Eldin. 2016, pp. 608–618.

[131] *Intelligent Cognitive Assistants: Workshop Summary and Recommendations*. Workshop Report. National Science Foundation, 2016.

[132] *Intelligent Cognitive Assistants: Workshop Summary and Research Needs*. Workshop Report. National Science Foundation, 2018.

[133] N.-T. Le and L. Wartschinski. "A cognitive assistant for improving human reasoning skills". In: *International Journal of Human-Computer Studies* 117 (2018), pp. 45–54. DOI: 10.1016/j.ijhcs.2018.02.005.

[134] S. Zhou and T. Bickmore. "A Virtual Counselor for Breast Cancer Genetic Counseling: Adaptive Pedagogy Leads to Greater Knowledge Gain". In: CHI '22. New Orleans, LA, USA: Association for Computing Machinery, 2022. DOI: 10.1145/3491102.3517553.

[135] F. Clarizia, F. Colace, M. De Santo, M. Lombardi, F. Pascale, and D. Santaniello. "A Context-Aware Chatbot for Tourist Destinations". In: *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2019, pp. 348–354. DOI: 10.1109/SITIS.2019.00063.

[136] E. Kimani, K. Rowan, D. McDuff, M. Czerwinski, and G. Mark. "A conversational agent in support of productivity and wellbeing at work". In: *2019 8th international conference on affective computing and intelligent interaction (ACII)*. IEEE. 2019, pp. 1–7. DOI: 10.1109/ACII.2019.8925488.

[137] S. Wellsandt, K. Hribernik, and K.-D. Thoben. “Anatomy of a Digital Assistant”. In: *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*. Ed. by A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski, and D. Romero. Vol. 633. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, 2021, pp. 321–330. ISBN: 978-3-030-85909-1. DOI: 10.1007/978-3-030-85910-7_34.

[138] F. G. Listl, J. Fischer, and M. Weyrich. “Towards a Simulation-based Conversational Assistant for the Operation and Engineering of Production Plants”. In: *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2021, pp. 1–4. ISBN: 978-1-7281-2989-1. DOI: 10.1109/ETFA45728.2021.9613681.

[139] M. Funk, T. Dingler, J. Cooper, and A. Schmidt. “Stop Helping Me - I’m Bored! Why Assembly Assistance Needs to Be Adaptive”. In: *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. UbiComp/ISWC’15 Adjunct. Osaka, Japan: Association for Computing Machinery, 2015, pp. 1269–1273. ISBN: 9781450335751. DOI: 10.1145/2800835.2807942.

[140] S. Wellsandt, M. Foosherian, K. Lepenioti, M. Fikardos, G. Mentzas, and K.-D. Thoben. “Supporting Data Analytics in Manufacturing with a Digital Assistant”. In: *Advances in Production Management Systems. Smart Manufacturing and Logistics Systems: Turning Ideas into Action*. Ed. by D. Y. Kim, G. von Cieminski, and D. Romero. Vol. 664. IFIP Advances in Information and Communication Technology. Cham: Springer Nature Switzerland, 2022, pp. 511–518. ISBN: 978-3-031-16410-1. DOI: 10.1007/978-3-031-16411-8_59.

[141] R. J. Rabelo, S. P. Zambiasi, and D. Romero. “Collaborative Softbots: Enhancing Operational Excellence in Systems of Cyber-Physical Systems”. In: *Collaborative Networks and Digital Transformation*. Ed. by L. M. Camarinha-Matos, H. Afsarmanesh, and D. Antonelli. Vol. 568. IFIP Advances in Information and Communication Technology. Springer Nature, 2019, pp. 55–68. ISBN: 978-3-030-28463-3. DOI: 10.1007/978-3-030-28464-0_6.

[142] B. Abner, R. J. Rabelo, S. P. Zambiasi, and D. Romero. “Production Management as-a-Service: A Softbot Approach”. In: *Advances in Production Management Systems. Towards Smart and Digital Manufacturing*. Ed. by B. Lalic, V. Majstorovic, U. Marjanovic, G. von Cieminski, and D. Romero. Vol. 592. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, 2020, pp. 19–30. ISBN: 978-3-030-57996-8. DOI: 10.1007/978-3-030-57997-5_3.

[143] L. Rodriguez, F. Quint, D. Gorecky, D. Romero, and H. R. Siller. “Developing a Mixed Reality Assistance System Based on Projection Mapping Technology for Manual Operations at Assembly Workstations”. en. In: *Procedia Computer Science*. 2015 International Conference Virtual and Augmented Reality in Education 75 (Jan. 2015), pp. 327–333. ISSN: 1877-0509. DOI: 10.1016/j.procs.2015.12.254.

6

[144] F. Belkadi, M. A. Dhuieb, J. V. Aguado, F. Laroche, A. Bernard, and F. Chinesta. “Intelligent assistant system as a context-aware decision-making support for the workers of the future”. en. In: *Computers & Industrial Engineering* 139 (Jan. 2020), p. 105732. ISSN: 0360-8352. DOI: 10.1016/j.cie.2019.02.046.

[145] S. Büttner, O. Sand, and C. Röcker. “Exploring Design Opportunities for Intelligent Worker Assistance: A New Approach Using Projetion-Based AR and a Novel Hand-Tracking Algorithm”. en. In: *Ambient Intelligence*. Ed. by A. Braun, R. Wichert, and A. Maña. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 33–45. ISBN: 978-3-319-56997-0. DOI: 10.1007/978-3-319-56997-0_3.

[146] K. Josifovska, E. Yigitbas, and G. Engels. “A Digital Twin-Based Multi-modal UI Adaptation Framework for Assistance Systems in Industry 4.0”. en. In: *Human-Computer Interaction. Design Practice in Contemporary Societies*. Ed. by M. Kurosu. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 398–409. ISBN: 978-3-030-22636-7. DOI: 10.1007/978-3-030-22636-7_30.

[147] F. Longo, L. Nicoletti, and A. Padovano. “Smart operators in industry 4.0: A human-centered approach to enhance operators’ capabilities and competencies within the new smart factory context”. In: *Computers & Industrial Engineering* 113 (2017), pp. 144–159. ISSN: 03608352. DOI: 10.1016/j.cie.2017.09.016.

[148] T. Kosch, M. Funk, A. Schmidt, and L. L. Chuang. “Identifying Cognitive Assistance with Mobile Electroencephalography: A Case Study with In-Situ Projections for Manual Assembly”. In: *Proceedings of the ACM on Human-Computer Interaction* 2.EICS (June 2018), 11:1–11:20. DOI: 10.1145/3229093.

[149] M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt. “A benchmark for interactive augmented reality instructions for assembly tasks”. In: *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*. MUM ’15. New York, NY, USA: Association for Computing Machinery, Nov. 2015, pp. 253–257. ISBN: 978-1-4503-3605-5. DOI: 10.1145/2836041.2836067.

[150] Y. Soliman and H. Vanharanta. “A Model for Capturing Tacit Knowledge in Enterprises”. en. In: *Advances in Human Factors, Business Management and Leadership*. Ed. by J. I. Kantola and S. Nazir. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2020, pp. 141–148. ISBN: 978-3-030-20154-8. DOI: 10.1007/978-3-030-20154-8_14.

[151] B. Dal, P. Tugwell, and R. Greatbanks. “Overall equipment effectiveness as a measure of operational improvement–a practical analysis”. In: *International Journal of Operations & Production Management* 20.12 (2000), pp. 1488–1502. DOI: 10.1108/01443570010355750.

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

[153] M. Schrepp, A. Hinderks, and J. Thomaschewski. “Construction of a Benchmark for the User Experience Questionnaire (UEQ)”. In: *International Journal of Interactive Multimedia and Artificial Intelligence* 4 (June 2017), pp. 40–44. DOI: 10.9781/ijimai.2017.445.

[154] A. Bangor, P. T. Kortum, and J. T. Miller. “An empirical evaluation of the system usability scale”. In: *Intl. Journal of Human–Computer Interaction* 24.6 (2008), pp. 574–594. DOI: 10.1080/10447310802205776.

[155] F. Brachten, T. Kissmer, and S. Stieglitz. “The acceptance of chatbots in an enterprise context – A survey study”. In: *International Journal of Information Management* 60 (2021), p. 102375. ISSN: 0268-4012. DOI: https://doi.org/10.1016/j.ijinfomgt.2021.102375.

[156] K. Savolainen. “User-Centred Design without Involving Users: A Longitudinal Case Study in a Human-Centred-Design–Mature Company”. In: *The Design Journal* 24.6 (2021), pp. 887–905. DOI: 10.1080/14606925.2021.1980267.

[157] Q. Roy, S. Berlioux, G. Casiez, and D. Vogel. “Typing Efficiency and Suggestion Accuracy Influence the Benefits and Adoption of Word Suggestions”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. Yokohama, Japan: Association for Computing Machinery, 2021. DOI: 10.1145/3411764.3445725.

6

[158] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz. “Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300484.

[159] M. Foosherian, S. Kernan Freire, E. Niforatos, K. A. Hribernik, and K.-D. Thoben. “Break, Repair, Learn, Break Less: Investigating User Preferences for Assignment of Divergent Phrasing Learning Burden in Human-Agent Interaction to Minimize Conversational Breakdowns”. In: MUM ’22. Lisbon, Portugal: Association for Computing Machinery, 2022, pp. 151–158. DOI: 10.1145/3568444.3568454.

[160] J. Lazar, J. H. Feng, and H. Hochheiser. “Chapter 11 - Analyzing qualitative data”. In: *Research Methods in Human Computer Interaction (Second Edition)*. Ed. by J. Lazar, J. H. Feng, and H. Hochheiser. Second Edition. Boston: Morgan Kaufmann, 2017, pp. 299–327. ISBN: 978-0-12-805390-4. DOI: https://doi.org/10.1016/B978-0-12-805390-4.00011-X.

[161] A. Bangor, P. Kortum, and J. Miller. “Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale”. In: *J. Usability Studies* 4.3 (May 2009), pp. 114–123.

[162] J. Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC, 2011.

[163] S. Holmes, A. Moorhead, R. Bond, H. Zheng, V. Coates, and M. Mctear. "Usability Testing of a Healthcare Chatbot: Can We Use Conventional Methods to Assess Conversational User Interfaces?" In: *Proceedings of the 31st European Conference on Cognitive Ergonomics*. ECCE '19. BELFAST, United Kingdom: Association for Computing Machinery, 2019, pp. 207–214. ISBN: 9781450371667. DOI: 10.1145/3335082.3335094.

[164] R. R. Holden and J. Passey. "Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance". In: *Personality and Individual Differences* 49.5 (2010), pp. 446–450. DOI: 10.1016/j.paid.2010.04.015.

[165] J. Alves, T. M. Lima, and P. D. Gaspar. "Is Industry 5.0 a Human-Centred Approach? A Systematic Review". In: *Processes* 11.1 (2023). ISSN: 2227-9717. DOI: 10.3390/pr11010193.

[166] G. May, M. Taisch, A. Bettoni, O. Maghazei, A. Matarazzo, and B. Stahl. "A New Human-centric Factory Model". In: *Procedia CIRP*. 12th Global Conference on Sustainable Manufacturing – Emerging Potentials 26 (2015), pp. 103–108. ISSN: 2212-8271. DOI: 10.1016/j.procir.2014.07.112.

[167] P. Fantini, M. Pinzone, and M. Taisch. "Placing the operator at the centre of Industry 4.0 design: Modelling and assessing human activities within cyber-physical systems". In: *Computers & Industrial Engineering* 139 (2020), p. 105058. ISSN: 0360-8352. DOI: https://doi.org/10.1016/j.cie.2018.01.025.

[168] "The Manufacturing Knowledge Repository - Consolidating Knowledge to Enable Holistic Process Knowledge Management in Manufacturing:" in: *Proceedings of the 16th International Conference on Enterprise Information Systems*. 16th International Conference on Enterprise Information Systems. Lisbon, Portugal: SCITEPRESS - Science and and Technology Publications, 2014, pp. 39–51. DOI: 10.5220/0004891200390051.

[169] L. Leoni, M. Ardolino, J. El Baz, G. Gueli, and A. Bacchetti. "The Mediating Role of Knowledge Management Processes in the Effective Use of Artificial Intelligence in Manufacturing Firms". In: *International Journal of Operations & Production Management* 42.13 (2022), pp. 411–437. ISSN: 0144-3577. DOI: 10.1108/IJOPM-05-2022-0282.

[170] S. Richter, L. Waizenegger, M. Steinhueser, and A. Richter. "Knowledge Management in the Dark: The Role of Shadow IT in Practices in Manufacturing". In: *International Journal of Knowledge Management (IJKM)* 15.2 (2019), pp. 1–19. ISSN: 1548-0666. DOI: 10.4018/IJKM.2019040101.

[171] S. Kernan Freire, C. Wang, S. Ruiz-Arenas, and E. Niforatos. "Tacit Knowledge Elicitation for Shop-Floor Workers with an Intelligent Assistant". In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA '23. Hamburg, Germany: Association for Computing Machinery, 2023. DOI: 10.1145/3544549.3585755.

[172] S. C. Feng, W. Z. Bernstein, J. Thomas Hedberg, and A. B. Feeney. "Toward Knowledge Management for Smart Manufacturing". In: *Journal of computing and information science in engineering* 17 3 (2017).

[173] G. Zuccon, B. Koopman, and R. Shaik. "ChatGPT Hallucinates when Attributing Answers". In: *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*. SIGIR-AP '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 46–51. DOI: 10.1145/3624918.3625329.

[174] P. K. R. Maddikunta, Q.-V. Pham, P. B, N. Deepa, K. Dev, T. R. Gadekallu, R. Ruby, and M. Liyanage. "Industry 5.0: A survey on enabling technologies and potential applications". In: *Journal of Industrial Information Integration* 26 (2022), p. 100257. ISSN: 2452-414X. DOI: https://doi.org/10.1016/j.jii.2021.100257.

[175] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang. "Industry 4.0 and Industry 5.0—Inception, conception and perception". In: *Journal of Manufacturing Systems* 61 (2021), pp. 530–535. ISSN: 0278-6125. DOI: https://doi.org/10.1016/j.jmsy.2021.10.006.

[176] A. Brückner, P. Hein, F. Hein-Pensel, J. Mayan, and M. Wölke. "Human-Centered HCI Practices Leading the Path to Industry 5.0: A Systematic Literature Review". In: *HCI International 2023 Posters*. Ed. by C. Stephanidis, M. Antona, S. Ntoa, and G. Salvendy. Cham: Springer Nature Switzerland, 2023, pp. 3–15. ISBN: 978-3-031-35989-7.

[177] S. Wellsandt, K. Hribernik, and K.-D. Thoben. "Anatomy of a Digital Assistant". In: *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*. Ed. by A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski, and D. Romero. Cham: Springer International Publishing, 2021, pp. 321–330. ISBN: 978-3-030-85910-7.

[178] T. Xie, Y. Wan, W. Huang, Z. Yin, Y. Liu, S. Wang, Q. Linghu, C. Kit, C. Grazian, W. Zhang, I. Razzak, and B. Hoex. *DARWIN Series: Domain Specific Large Language Models for Natural Science*. 2023. DOI: 10.48550/arXiv.2308.13565. preprint.

[179] C. Wen, X. Sun, S. Zhao, X. Fang, L. Chen, and W. Zou. *ChatHome: Development and Evaluation of a Domain-Specific Language Model for Home Renovation*. 2023. DOI: 10.48550/arXiv.2307.15290. preprint.

[180] J. Zhang, Y. Chen, N. Niu, and C. Liu. "A Preliminary Evaluation of ChatGPT in Requirements Information Retrieval". In: (Apr. 2023).

[181] J. Wei, X. Wang, D. Schuurmans, M. Bosma, b. ichter brian, F. Xia, E. Chi, Q. V. Le, and D. Zhou. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models". In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. Vol. 35. Curran Associates, Inc., 2022, pp. 24824–24837.

[182] S. J. Semnani, V. Z. Yao, H. C. Zhang, and M. S. Lam. *WikiChat: A Few-Shot LLM-Based Chatbot Grounded with Wikipedia*. 2023. arXiv: 2305.14292 [cs.CL]. preprint.

6

[183] R. Tang, X. Han, X. Jiang, and X. Hu. *Does Synthetic Data Generation of LLMs Help Clinical Text Mining?* 2023. arXiv: 2303.04360 [cs.CL]. preprint.

[184] O. Nov, N. Singh, and D. Mann. *Putting ChatGPT's Medical Advice to the (Turing) Test.* 2023. arXiv: 2301.10035. preprint.

[185] K. Jeblick, B. Schachtner, J. Dexl, A. Mittermeier, A. T. Stüber, J. Topalis, T. Weber, P. Wesp, B. Sabel, J. Ricke, and M. Ingrisch. *ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports.* 2022. arXiv: 2212.14882 [cs.CL]. preprint.

[186] D. Trautmann, A. Petrova, and F. Schilder. *Legal Prompt Engineering for Multilingual Legal Judgement Prediction.* 2022. arXiv: 2212.02199 [cs.CL]. preprint.

[187] H. Alkaissi and S. I. McFarlane. "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing". In: *Cureus* 15.2 (2023), e35179. ISSN: 2168-8184. DOI: 10.7759/cureus.35179.

[188] C. Wei, Y.-C. Wang, B. Wang, and C. .-. J. Kuo. "An Overview on Language Models: Recent Developments and Outlook". In: (Mar. 2023). DOI: https://doi.org/10.48550/arXiv.2303.05759. preprint.

[189] F. F. Xu, U. Alon, G. Neubig, V. J. Hellendoorn, and V. J. Hel. "A systematic evaluation of large language models of code". In: (June 2022), pp. 1–10. DOI: 10.1145/3520312.3534862.

[190] J. Jang, S. Ye, C. Lee, S. Yang, J. Shin, J. Han, G. Kim, and M. Seo. "TemporalWiki: A Lifelong Benchmark for Training and Evaluating Ever-Evolving Language Models". In: (2022).

[191] B. C. Kwon and N. Mihindukulasooriya. "An Empirical Study on Pseudo-log-likelihood Bias Measures for Masked Language Models Using Paraphrased Sentences". In: *TrustNLP 2022 - 2nd Workshop on Trustworthy Natural Language Processing, Proceedings of the Workshop* (2022), pp. 74–79. DOI: 10.18653/V1/2022.TRUSTNLP-1.7.

[192] R. Zellers, A. Holtzman, Y. Bisk, A. Farhadi, and Y. Choi. "HellaSwag: Can a Machine Really Finish Your Sentence?" In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* Ed. by A. Korhonen, D. Traum, and L. Màrquez. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 4791–4800. DOI: 10.18653/v1/P19-1472.

[193] M. Joshi, E. Choi, D. S. Weld, and L. Zettlemoyer. *TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension.* 2017. DOI: 10.48550/arXiv.1705.03551. arXiv: 1705.03551. preprint.

[194] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt. *Measuring Massive Multitask Language Understanding.* 2021. DOI: 10.48550/arXiv.2009.03300. arXiv: 2009.03300. preprint.

[195] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, P. Payne, M. Seneviratne, P. Gamble, C. Kelly, A. Babiker, N. Schärli, A. Chowdhery, P. Mansfield, D. Demner-Fushman, B. Agüera y Arcas, D. Webster, G. S. Corrado, Y. Matias, K. Chou, J. Gottweis, N. Tomasev, Y. Liu, A. Rajkomar, J. Barral, C. Semturs, A. Karthikesalingam, and V. Natarajan. "Large Language Models Encode Clinical Knowledge". In: *Nature* 620.7972 (Aug. 2023), pp. 172–180. DOI: 10.1038/s41586-023-06291-2.

[196] Q. Xie, W. Han, X. Zhang, Y. Lai, M. Peng, A. Lopez-Lira, and J. Huang. "PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance". In: (June 2023).

[197] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng. "Code as Policies: Language Model Programs for Embodied Control". In: (Sept. 2022).

[198] S. Badini, S. Regondi, E. Frontoni, and R. Pugliese. "Assessing the capabilities of ChatGPT to improve additive manufacturing troubleshooting". In: *Advanced Industrial and Engineering Polymer Research* (Mar. 2023). ISSN: 2542-5048. DOI: 10.1016/J.AIEPR.2023.03.003.

[199] A. Abid, A. Abdalla, A. Abid, D. Khan, A. Alfozan, and J. Zou. *Gradio: Hassle-free sharing and testing of ML models in the wild.* June 2019. DOI: 10.48550/arXiv.1906.02569. preprint.

[200] J. Liu. *LlamaIndex.* Nov. 2022. DOI: 10.5281/zenodo.1234.

[201] O. Serrat. "The five whys technique". In: *Knowledge solutions: Tools, methods, and approaches to drive organizational performance* (2017), pp. 307–310.

[202] G. Guest, K. M. MacQueen, and E. E. Namey. *Applied thematic analysis.* sage publications, 2011.

[203] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer. *QLoRA: Efficient Finetuning of Quantized LLMs.* 2023. arXiv: 2305.14314 [cs.LG]. preprint.

[204] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample. *LLaMA: Open and Efficient Foundation Language Models.* 2023. DOI: 10.48550/arXiv.2302.13971. arXiv: 2302.13971 [cs.CL]. preprint.

[205] A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. S. Chaplot, D. de las Casas, E. B. Hanna, F. Bressand, G. Lengyel, G. Bour, G. Lample, L. R. Lavaud, L. Saulnier, M.-A. Lachaux, P. Stock, S. Subramanian, S. Yang, S. Antoniak, T. L. Scao, T. Gervet, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed. *Mixtral of Experts.* 2024. arXiv: 2401.04088 [cs.LG]. preprint.

[206] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton,

J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, and T. Scialom. *Llama 2: Open Foundation and Fine-Tuned Chat Models*. 2023. arXiv: 2307.09288 [cs.CL]. preprint.

[207] D. Mahan, R. Carlow, L. Castricato, N. Cooper, and C. Laforte. *Stable Beluga models*. 2023.

[208] OpenAI. *GPT-4 Technical Report*. 2023. arXiv: 2303.08774 [cs.CL]. preprint.

[209] M. Eswaran and M. V. A. R. Bahubalendruni. "Challenges and opportunities on AR/VR technologies for manufacturing systems in the context of industry 4.0: A state of the art review". In: *Journal of Manufacturing Systems* 65 (2022), pp. 260–278. ISSN: 0278-6125. DOI: https://doi.org/10.1016/j.jmsy.2022.09.016.

[210] E. Rauch, P. Dallasega, and M. Unterhofer. "Requirements and Barriers for Introducing Smart Manufacturing in Small and Medium-Sized Enterprises". In: *IEEE Engineering Management Review* 47.3 (2019), pp. 87–94. DOI: 10.1109/EMR.2019.2931564.

[211] P. K. R. Maddikunta, Q.-V. Pham, P. B, N. Deepa, K. Dev, T. R. Gadekallu, R. Ruby, and M. Liyanage. "Industry 5.0: A Survey on Enabling Technologies and Potential Applications". In: *Journal of Industrial Information Integration* 26 (Mar. 1, 2022), p. 100257. DOI: 10.1016/j.jii.2021.100257.

[212] A. Adel. "Future of Industry 5.0 in Society: Human-Centric Solutions, Challenges and Prospective Research Areas". In: *Journal of Cloud Computing* 11.1 (Sept. 8, 2022), p. 40. DOI: 10.1186/s13677-022-00314-5.

[213] E. Cheon, E. Schneiders, and M. B. Skov. "Working with Bounded Collaboration: A Qualitative Study on How Collaboration Is Co-constructed around Collaborative Robots in Industry". In: *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW2 Nov. 11, 2022), 369:1–369:34. DOI: 10.1145/3555094.

[214] N. M. Dixon. *Common knowledge: How companies thrive by sharing what they know*. Harvard Business School Press, 2000.

[215] S. Trussler. "The rules of the game." In: *Journal of business strategy* 19.1 (1998), pp. 16–19.

[216] T. Davenport and L. Prusak. *Working knowledge: How organizations manage what they know*. Harvard Business Press, 1998.

[217] R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg. "DENDRAL: A case study of the first expert system for scientific hypothesis formation". In: *Artificial Intelligence* 61.2 (1993), pp. 209–261. ISSN: 0004-3702. DOI: https://doi.org/10.1016/0004-3702(93)90068-M.

[218] E. H. Shortliffe, R. Davis, S. G. Axline, B. G. Buchanan, C. C. Green, and S. N. Cohen. "Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system". In: *Computers and Biomedical Research* 8.4 (1975), pp. 303–320. ISSN: 0010-4809. DOI: https://doi.org/10.1016/0010-4809(75)90009-9.

[219] R. High. “The era of cognitive systems: An inside look at IBM Watson and how it works”. In: *IBM Corporation, Redbooks* 1 (2012), p. 16.

[220] M. Breque, L. de Nul, and A. Petridis. *Industry 5.0 - Towards a Sustainable, Human-Centric and Resilient European Industry - European Commission*. European Commission, 2021.

[221] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang. “Industry 4.0 and Industry 5.0—Inception, Conception and Perception”. In: *Journal of Manufacturing Systems* 61 (Oct. 1, 2021), pp. 530–535. ISSN: 0278-6125. DOI: 10.1016/j.jmsy.2021.10.006.

[222] H. Mucha, C. Roecker, T. Ludwig, C. Ogonowski, M. Stein, S. Robert, L. Galla, M. Hill, and V. Wulf. “The Industrial Internet of Things: New Perspectives on HCI and CSCW within Industry Settings”. In: *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. CSCW ’18 Companion. New York, NY, USA: Association for Computing Machinery, Oct. 30, 2018, pp. 393–400. ISBN: 978-1-4503-6018-0. DOI: 10.1145/3272973.3273009.

[223] T. Weinert, M. Billert, M. T. de Gafenco, A. Janson, and J. M. Leimeister. “Designing a Co-Creation System for the Development of Work-Process-Related Learning Material in Manufacturing”. In: *Comput. Supported Coop. Work* 32.1 (Jan. 2022), pp. 5–53. ISSN: 0925-9724. DOI: 10.1007/s10606-021-09420-5.

[224] K. Boehner, J. Vertesi, P. Sengers, and P. Dourish. “How HCI interprets the probes”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2007, pp. 1077–1086. DOI: 10.1145/1240624.1240789.

[225] R. Martinez-Maldonado, V. Echeverria, G. Fernandez-Nieto, L. Yan, L. Zhao, R. Alfredo, X. Li, S. Dix, H. Jaggard, R. Wotherspoon, A. Osborne, S. B. Shum, and D. Gašević. “Lessons Learnt from a Multimodal Learning Analytics Deployment In-the-wild”. In: *ACM Transactions on Computer-Human Interaction* (Sept. 2023). Just Accepted. ISSN: 1073-0516. DOI: 10.1145/3622784.

[226] D. G. Schniederjans, C. Curado, and M. Khalajhedayati. “Supply Chain Digitisation Trends: An Integration of Knowledge Management”. In: *International Journal of Production Economics* 220 (Feb. 1, 2020), p. 107439. ISSN: 0925-5273. DOI: 10.1016/j.ijpe.2019.07.012.

[227] A. P. Lista Rossetti, G. Luz Tortorella, M. Bouzon, S. Gao, and T. K. Chan. “Identifying Industry 4.0 Technologies Enablers for Knowledge Management – a Scoping Review”. In: *The TQM Journal* 36.1 (Jan. 1, 2023), pp. 340–360. ISSN: 1754-2731. DOI: 10.1108/TQM-05-2022-0173.

[228] M. Gerbracht, M. Krüger, D. De Castro Leal, P. Tolmie, and V. Wulf. “"We Are No Luddites!" - CSCW, Co-Determination and Digital Transformation in Germany”. In: *Proceedings of the ACM on Human-Computer Interaction* 8 (CSCW1 Apr. 26, 2024), 153:1–153:23. DOI: 10.1145/3637430.

[229] F. D. Davis. “Perceived usefulness, perceived ease of use, and user acceptance of information technology”. In: *MIS quarterly* (1989), pp. 319–340.

6

[230] C.-L. Yang, C. W. (Yuan, and H.-C. Wang. “When Knowledge Network Is Social Network: Understanding Collaborative Knowledge Transfer in Workplace”. In: *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW Nov. 7, 2019), 164:1–164:23. DOI: 10.1145/3359266.

[231] A. Crabtree, P. Tolmie, and W. Knight. “Repacking ‘privacy’ for a networked world”. In: *Computer Supported Cooperative Work (CSCW)* 26 (2017), pp. 453–488. DOI: 10.1007/s10606-017-9276-y.

[232] R. Zhang, N. J. McNeese, G. Freeman, and G. Musick. “"An Ideal Human": Expectations of AI Teammates in Human-AI Teaming”. In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW3 (Jan. 2021). DOI: 10.1145/3432945.

[233] W. K. Edwards, M. W. Newman, and E. S. Poole. “The Infrastructure Problem in HCI”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’10. Atlanta, Georgia, USA: Association for Computing Machinery, 2010, pp. 423–432. ISBN: 9781605589299. DOI: 10.1145/1753326.1753390.

[234] S. Brahe and K. Schmidt. “The Story of a Working Workflow Management System”. In: *Proceedings of the 2007 ACM International Conference on Supporting Group Work*. GROUP ’07. New York, NY, USA: Association for Computing Machinery, Nov. 4, 2007, pp. 249–258. ISBN: 978-1-59593-845-9. DOI: 10.1145/1316624.1316661.

[235] B. D. Rosso, K. H. Dekas, and A. Wrzesniewski. “On the Meaning of Work: A Theoretical Integration and Review”. In: *Research in Organizational Behavior* 30 (Jan. 1, 2010), pp. 91–127. ISSN: 0191-3085. DOI: 10.1016/j.riob.2010.09.001.

[236] V. Das Swain, L. Gao, A. Mondal, G. D. Abowd, and M. De Choudhury. “Sensible and Sensitive AI for Worker Wellbeing: Factors That Inform Adoption and Resistance for Information Workers”. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI ’24. New York, NY, USA: Association for Computing Machinery, May 11, 2024, pp. 1–30. DOI: 10.1145/3613904.3642716.

[237] S. E. Fox, V. Khovanskaya, C. Crivellaro, N. Salehi, L. Dombrowski, C. Kulkarni, L. Irani, and J. Forlizzi. “Worker-Centered Design: Expanding HCI Methods for Supporting Labor”. In: CHI EA ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–8. ISBN: 9781450368193. DOI: 10.1145/3334480.3375157.

[238] T. Meneweger, D. Wurhofer, V. Fuchsberger, and M. Tscheligi. “Working Together with Industrial Robots: Experiencing Robots in a Production Environment”. In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). Aug. 2015, pp. 833–838. DOI: 10.1109/ROMAN.2015.7333641.

[239] Z. Buçinca, M. B. Malaya, and K. Z. Gajos. “To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making”. In: *Proc. ACM Hum.-Comput. Interact.* 5.CSCW1 (Apr. 2021). DOI: 10.1145/3449287.

[240] M. A. Dhuieb, F. Laroche, and A. Bernard. “Context-awareness: A Key Enabler for Ubiquitous Access to Manufacturing Knowledge”. In: *Procedia CIRP* 41 (2016). Research and Innovation in Manufacturing: Key Enabling Technologies for the Factories of the Future - Proceedings of the 48th CIRP Conference on Manufacturing Systems, pp. 484–489. ISSN: 2212-8271. DOI: https://doi.org/10.1016/j.procir.2015.12.027.

[241] N. A. J. Cornelissen, R. J. M. van Eerdt, H. K. Schraffenberger, and W. F. G. Haselager. “Reflection Machines: Increasing Meaningful Human Control over Decision Support Systems”. In: *Ethics and Information Technology* 24.2 (Mar. 21, 2022), p. 19. ISSN: 1572-8439. DOI: 10.1007/s10676-022-09645-y.

[242] T. Ha and S. Kim. “Improving Trust in AI with Mitigating Confirmation Bias: Effects of Explanation Type and Debiasing Strategy for Decision-Making with Explainable AI”. In: *International Journal of Human–Computer Interaction* 0.0 (2023), pp. 1–12. DOI: 10.1080/10447318.2023.2285640. eprint: https://doi.org/10.1080/10447318.2023.2285640.

[243] A. Kluge and N. Gronau. “Intentional Forgetting in Organizations: The Importance of Eliminating Retrieval Cues for Implementing New Routines”. In: *Frontiers in Psychology* 9 (2018). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2018.00051.

[244] T. L. Johnson, S. R. Fletcher, W. Baker, and R. L. Charles. “How and why we need to capture tacit knowledge in manufacturing: Case studies of visual inspection”. In: *Applied ergonomics* 74 (2019), pp. 1–9. DOI: 10.1016/j.apergo.2018.07.016.

[245] S. M. R. Naqvi, M. Ghufran, S. Meraghni, C. Varnier, J.-M. Nicod, and N. Zerhouni. “Human knowledge centered maintenance decision support in digital twin environment”. In: *Journal of Manufacturing Systems* 65 (2022), pp. 528–537. ISSN: 0278-6125. DOI: https://doi.org/10.1016/j.jmsy.2022.10.003.

[246] P. Heikkilä, A. Honka, S. Mach, F. Schmalfuß, E. Kaasinen, and K. Väänänen. “Quantified Factory Worker - Expert Evaluation and Ethical Considerations of Wearable Self-tracking Devices”. In: *Proceedings of the 22nd International Academic Mindtrek Conference*. Mindtrek ’18. Tampere, Finland: Association for Computing Machinery, 2018, pp. 202–211. ISBN: 9781450365895. DOI: 10.1145/3275116.3275119.

[247] I. Bisio, C. Garibotto, A. Grattarola, F. Lavagetto, and A. Sciarrone. “Exploiting Context-Aware Capabilities over the Internet of Things for Industry 4.0 Applications”. In: *IEEE Network* 32.3 (May 2018), pp. 101–107. ISSN: 1558-156X. DOI: 10.1109/MNET.2018.1700355.

[248] C.-H. Li, S.-F. Yeh, T.-J. Chang, M.-H. Tsai, K. Chen, and Y.-J. Chang. “A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–12. ISBN: 9781450367080. DOI: 10.1145/3313831.3376209.

6

[249] J. Hill, W. Randolph Ford, and I. G. Farreras. “Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations”. In: *Computers in Human Behavior* 49 (2015), pp. 245–250. ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2015.02.026.

[250] S.-F. Yeh, M.-H. Wu, T.-Y. Chen, Y.-C. Lin, X. Chang, Y.-H. Chiang, and Y.-J. Chang. “How to Guide Task-oriented Chatbot Users, and When: A Mixed-methods Study of Combinations of Chatbot Guidance Types and Timings”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI ’22. New York, NY, USA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3501941.

[251] H. Flatt, N. Koch, C. Röcker, A. Günter, and J. Jasperneite. “A context-aware assistance system for maintenance applications in smart factories based on augmented reality and indoor localization”. In: *2015 IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA)*. 2015, pp. 1–4. DOI: 10.1109/ETFA.2015.7301586.

[252] E. Bottani, F. Longo, L. Nicoletti, A. Padovano, G. P. C. Tancredi, L. Tebaldi, M. Vetrano, and G. Vignali. “Wearable and interactive mixed reality solutions for fault diagnosis and assistance in manufacturing systems: Implementation and testing in an aseptic bottling line”. In: *Computers in Industry* 128 (2021), p. 103429. ISSN: 0166-3615. DOI: https://doi.org/10.1016/j.compind.2021.103429.

[253] A. Følstad and P. B. Brandtzaeg. “Users’ Experiences with Chatbots: Findings from a Questionnaire Study”. In: *Quality and User Experience* 5.1 (Apr. 11, 2020), p. 3. ISSN: 2366-0147. DOI: 10.1007/s41233-020-00033-2.

[254] H. Vasconcelos, M. Jörke, M. Grunde-McLaughlin, T. Gerstenberg, M. S. Bernstein, and R. Krishna. “Explanations Can Reduce Overreliance on AI Systems During Decision-Making”. In: 7.CSCW1 (Apr. 2023). DOI: 10.1145/3579605.

[255] S. Cao and C.-M. Huang. “Understanding User Reliance on AI in Assisted Decision-Making”. In: 6.CSCW2 (Nov. 2022). DOI: 10.1145/3555572.

[256] C. Lindblom. “The science of “muddling through””. In: *Public administration review*. 1959.

[257] A. Raj, G. Dwivedi, A. Sharma, A. B. Lopes de Sousa Jabbour, and S. Rajak. “Barriers to the adoption of industry 4.0 technologies in the manufacturing sector: An inter-country comparative perspective”. In: *International Journal of Production Economics* 224 (2020), p. 107546. ISSN: 0925-5273. DOI: https://doi.org/10.1016/j.ijpe.2019.107546.

[258] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang. *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. 2023. DOI: 10.48550/arXiv.2303.12712. arXiv: 2303.12712 [cs.CL]. preprint.

[259] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, S. Zhong, B. Yin, and X. Hu. "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond". In: *ACM Transactions on Knowledge Discovery from Data* (Feb. 28, 2024). ISSN: 1556-4681. DOI: 10.1145/3649506.

[260] M. Besta, N. Blach, A. Kubicek, R. Gerstenberger, M. Podstawski, L. Gianinazzi, J. Gajda, T. Lehmann, H. Niewiadomski, P. Nyczyk, and T. Hoefler. *Graph of Thoughts: Solving Elaborate Problems with Large Language Models*. Feb. 6, 2024. DOI: 10.48550/arXiv.2308.09687. arXiv: 2308.09687 [cs]. preprint.

[261] Y. Zhang, J. Yang, Y. Yuan, and A. C.-C. Yao. *Cumulative Reasoning with Large Language Models*. Dec. 1, 2023. DOI: 10.48550/arXiv.2308.04371. arXiv: 2308.04371 [cs]. preprint.

[262] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, and K. Narasimhan. *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*. Dec. 3, 2023. DOI: 10.48550/arXiv.2305.10601. arXiv: 2305.10601 [cs]. preprint.

[263] Y. Zhang, Y. Yuan, and A. C.-C. Yao. *Meta Prompting for AGI Systems*. Jan. 31, 2024. DOI: 10.48550/arXiv.2311.11482. arXiv: 2311.11482 [cs]. preprint.

[264] P. Zhou, J. Pujara, X. Ren, X. Chen, H.-T. Cheng, Q. V. Le, E. H. Chi, D. Zhou, S. Mishra, and H. S. Zheng. *Self-Discover: Large Language Models Self-Compose Reasoning Structures*. Feb. 5, 2024. arXiv: 2402.03620 [cs].

[265] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang. *Lost in the Middle: How Language Models Use Long Contexts*. Nov. 20, 2023. DOI: 10.48550/arXiv.2307.03172. arXiv: 2307.03172 [cs]. preprint.

[266] H. Jiang, Q. Wu, X. Luo, D. Li, C.-Y. Lin, Y. Yang, and L. Qiu. *LongLLMLingua: Accelerating and Enhancing LLMs in Long Context Scenarios via Prompt Compression*. Oct. 10, 2023. DOI: 10.48550/arXiv.2310.06839. arXiv: 2310.06839 [cs]. preprint.

[267] M. Steyvers and A. Kumar. "Three Challenges for AI-Assisted Decision-Making". In: *Perspectives on Psychological Science* 0.0 (2023). DOI: 10.1177/17456916231181102.

[268] P. P. Ray. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope". In: *Internet of Things and Cyber-Physical Systems* 3 (2023), pp. 121–154. ISSN: 2667-3452. DOI: 10.1016/j.iotcps.2023.04.003.

[269] B. Shneiderman. "Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems". In: *ACM Transactions on Interactive Intelligent Systems* 10.4 (Oct. 16, 2020), 26:1–26:31. ISSN: 2160-6455. DOI: 10.1145/3419764.

[270] Y. Deng, L. Liao, L. Chen, H. Wang, W. Lei, and T.-S. Chua. *Prompting and Evaluating Large Language Models for Proactive Dialogues: Clarification, Target-guided, and Non-collaboration*. 2023. DOI: 10.48550/arXiv.2305.13626. arXiv: 2305.13626 [cs.CL]. preprint.

[271] S. Lin, J. Warner, J. D. Zamfirescu-Pereira, M. G. Lee, S. Jain, M. X. Huang, P. Lertvittayakumjorn, S. Cai, S. Zhai, B. Hartmann, and C. Liu. *Rambler: Supporting Writing With Speech via LLM-Assisted Gist Manipulation*. 2024. DOI: 10.48550/arXiv.2401.10838. preprint.

[272] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang. "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development". In: *IEEE/-CAA Journal of Automatica Sinica* 10.5 (2023), pp. 1122–1136. DOI: 10.1109/JAS.2023.123618.

[273] J. Retkowsky, E. Hafermalz, and M. Huysman. "Managing a ChatGPT-empowered workforce: Understanding its affordances and side effects". In: *Business Horizons* (2024). ISSN: 0007-6813. DOI: https://doi.org/10.1016/j.bushor.2024.04.009.

[274] N. Sakib, F. I. Anik, and L. Li. "ChatGPT in IT Education Ecosystem: Unraveling Long-Term Impacts on Job Market, Student Learning, and Ethical Practices". In: *Proceedings of the 24th Annual Conference on Information Technology Education*. SIGITE '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 73–78. DOI: 10.1145/3585059.3611447.

[275] S. Saniuk, D. Caganova, and A. Saniuk. "Knowledge and Skills of Industrial Employees and Managerial Staff for the Industry 4.0 Implementation". In: *Mobile Networks and Applications* 28.1 (Feb. 1, 2023), pp. 220–230. ISSN: 1572-8153. DOI: 10.1007/s11036-021-01788-4.

[276] C. F. Fey and P. Furu. "Top Management Incentive Compensation and Knowledge Sharing in Multinational Corporations". In: *Strategic Management Journal* 29.12 (2008), pp. 1301–1323. ISSN: 1097-0266. DOI: 10.1002/smj.712.

[277] L.-A. Ho and T.-H. Kuo. "How System Quality and Incentive Affect Knowledge Sharing". In: *Industrial Management & Data Systems* 113.7 (Jan. 1, 2013), pp. 1048–1063. ISSN: 0263-5577. DOI: 10.1108/IMDS-01-2013-0015.

[278] L. Hu and A. E. Randel. "Knowledge Sharing in Teams: Social Capital, Extrinsic Incentives, and Team Innovation". In: *Group & Organization Management* 39.2 (Apr. 1, 2014), pp. 213–243. ISSN: 1059-6011. DOI: 10.1177/1059601114520969.

[279] H. Lyu and Z. (Zhang. "Incentives for Knowledge Sharing: Impact of Organisational Culture and Information Technology". In: *Enterprise Information Systems* 11.9 (Oct. 21, 2017), pp. 1416–1435. ISSN: 1751-7575. DOI: 10.1080/17517575.2016.1273393.

[280] E. Z. Taylor. "The Effect of Incentives on Knowledge Sharing in Computer-Mediated Communication: An Experimental Investigation". In: *Journal of Information Systems* 20.1 (Mar. 1, 2006), pp. 103–116. ISSN: 0888-7985. DOI: 10.2308/jis.2006.20.1.103.

[281] Z. Zhang. "Organizational Culture and Knowledge Sharing: Design of Incentives and Business Processes". In: *Business Process Management Journal* 24.2 (Jan. 1, 2018), pp. 384–399. ISSN: 1463-7154. DOI: 10.1108/BPMJ-08-2015-0119.

6

[282] L. Razmerita, K. Kirchner, and P. Nielsen. "What Factors Influence Knowledge Sharing in Organizations? A Social Dilemma Perspective of Social Media Communication". In: *Journal of Knowledge Management* 20.6 (Jan. 1, 2016), pp. 1225–1246. ISSN: 1367-3270. DOI: 10.1108/JKM-03-2016-0112.

[283] C. Wang, S. Kernan Freire, M. Zhang, J. Wei, J. Goncalves, V. Kostakos, Z. Sarsenbayeva, C. Schneegass, A. Bozzon, and E. Niforatos. "Safeguarding Crowdsourcing Surveys from ChatGPT with Prompt Injection". In: *Proceedings of the ACM on Human-Computer Interaction* (CSCW 2025). submitted.

[284] H. Torkamaan, S. Steinert, M. Soledad Pera, O. Kudina, S. Kernan Freire, H. Verma, S. Kelly, M.-T. Sekwenz, J. Yang, K. van Nunen, M. Warnier, and O. Oviedo-Trespalacios. "Challenges and Future Directions for Integration of Large Language Models into Socio-technical System". In: *Behaviour & Information Technology* (2024). submitted.

[285] S. Kernan Freire, C. Wang, and E. Niforatos. *Chatbots in Knowledge-Intensive Contexts: Comparing Intent and LLM-Based Systems*. 2024. DOI: 10.48550/arXiv.2402.04955. preprint.

[286] S. Kernan Freire, M. M. van Mol, C. Schol, and E. Ö. Vieira. "Enhancing ICU Patient Recovery: Using LLMs to Assist Nurses in Diary Writing". In: *ACM Transactions on Computing for Healthcare* (2024). submitted.

[287] S. Kernan Freire, M. Foosherian, C. Wang, and E. Niforatos. "Harnessing Large Language Models for Cognitive Assistants in Factories". In: *Proceedings of the 5th International Conference on Conversational User Interfaces*. CUI '23. Eindhoven, Netherlands: Association for Computing Machinery, 2023. DOI: 10.1145/3571884.3604313.

[288] S. Kernan Freire, J.-Y. Jung, C. Wang, E. Niforatos, and A. Bozzon. "How Emoji and Explanations Influence Adherence to AI Recommendations". In: *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. IVA '23. Würzburg, Germany: Association for Computing Machinery, 2023. DOI: 10.1145/3570945.3607300.

Acknowledgements

I want to thank the following people for their invaluable support and contributions to this dissertation.

Firstly, thank you to my supervisors. Evan, thank you for your guidance, honesty, and insightful feedback throughout this journey. Alessandro, for your constructive critiques and support. Zoltan and Doris, thank you for guiding me as I began this journey.

I'd like to thank the rest of the IDE COALA team, Evan, Chaofan, Santiago, Doris, Soude, Sarath, Adhi, Carlo, Matthijs, and Zoltan, for their discussions and contributions. Also, I'd like to extend my appreciation to the engaging collaborations with the COALA consortium members, especially Stefan, Mina, Jan and Petros, and the industrial partners that opened their factories to us.

My KInD colleagues and office mates, thank you for your camaraderie and assistance. Our lunch breaks, coffees, and walks were always fun!

My parents, Margaret and Filipe, and my sister Eva, thank you for your understanding and encouragement.

Lastly, I feel immense gratitude to my wife, Kim, and kids, Julie and Liam, for listening, encouraging, and joining me on the journey, oftentimes physically, as I attended conferences and other events around Europe.

Thank you all for your contributions and support. This dissertation would not have been possible without you.

About the author

Samuel Kernan Freire was born on May 28th, 1992, in London, United Kingdom, to Irish and Portuguese parents.

In 2003, after moving around Europe, his family settled near Leiden, The Netherlands. There, he attended The Rijnlands Lyceum until 2010, having followed the International Baccalaureate program. From 2010 to 2014, he studied Aerospace Engineering at the Delft University of Technology but did not complete the degree. After working for a year, he returned to the Delft University of Technology from 2015 to 2020, attaining a Bachelor of Science in Industrial Design Engineering and a Master of Science in Integrated Product Design. During this period, Samuel worked as a product design engineer in various contexts: as a freelancer, at a startup, and a design agency.

Samuel conducted this research from 2020 to 2024 at the Delft University of Technology in the Knowledge and Intelligence Design group. He lives in Delft, The Netherlands, with his wife, Kim, and two children, Julie and Liam.

List of Publications

1. Full Conference, Journal and Magazine Articles

- S. Kernan Freire, T. He, C. Wang, E. Niforatos, and A. Bozzon. “Operators’ Perspectives on Conversational AI for Knowledge Sharing: Challenges, Risks and Impact on Work”. In: *Proceedings of the ACM on Human-Computer Interaction* (CSCW 2025). submitted.
- C. Wang, S. Kernan Freire, M. Zhang, J. Wei, J. Goncalves, V. Kostakos, Z. Sarsenbayeva, C. Schneegass, A. Bozzon, and E. Niforatos. “Safeguarding Crowdsourcing Surveys from ChatGPT with Prompt Injection”. In: *Proceedings of the ACM on Human-Computer Interaction* (CSCW 2025). submitted.
- H. Torkamaan, S. Steinert, M. Soledad Pera, O. Kudina, S. Kernan Freire, H. Verma, S. Kelly, M.-T. Sekwenz, J. Yang, K. van Nunen, M. Warnier, and O. Oviedo-Trespalacios. “Challenges and Future Directions for Integration of Large Language Models into Socio-technical System”. In: *Behaviour & Information Technology* (2024). submitted.
- S. Kernan Freire, C. Wang, M. Foosherian, S. Wellsandt, S. Ruiz-Arenas, and E. Niforatos. “Knowledge sharing in manufacturing using LLM-powered tools: user study and model benchmarking”. In: *Frontiers in Artificial Intelligence* 7 (2024). ISSN: 2624-8212. DOI: 10.3389/frai.2024.1293084
- S. Kernan Freire, E. Niforatos, C. Wang, S. Ruiz-Arenas, M. Foosherian, S. Wellsandt, and A. Bozzon. “Lessons Learned from Designing and Evaluating CLAICA: A Continuously Learning AI Cognitive Assistant”. In: *Proceedings of the 28th International Conference on Intelligent User Interfaces*. IUI ’23. Sydney, NSW, Australia: Association for Computing Machinery, 2023, pp. 553–568. DOI: 10.1145/3581641.3584042
- S. Kernan Freire, S. S. Panicker, S. Ruiz-Arenas, Z. Rusák, and E. Niforatos. “A Cognitive Assistant for Operators: AI-Powered Knowledge Sharing on Complex Systems”. In: *IEEE Pervasive Computing* (2022), pp. 1–9. DOI: 10.1109/MPRV.2022.3218600

2. Late Breaking Work, Posters and Short Papers

- S. Kernan Freire, C. Wang, and E. Niforatos. *Chatbots in Knowledge-Intensive Contexts: Comparing Intent and LLM-Based Systems*. 2024. DOI: 10.48550/arXiv.2402.04955. preprint

- S. Kernan Freire, M. M. van Mol, C. Schol, and E. Ö. Vieira. "Enhancing ICU Patient Recovery: Using LLMs to Assist Nurses in Diary Writing". In: *ACM Transactions on Computing for Healthcare* (2024). submitted.

- S. Kernan Freire, M. Foosherian, C. Wang, and E. Niforatos. "Harnessing Large Language Models for Cognitive Assistants in Factories". In: *Proceedings of the 5th International Conference on Conversational User Interfaces*. CUI '23. Eindhoven, Netherlands: Association for Computing Machinery, 2023. DOI: 10.1145/3571884.3604313

- S. Kernan Freire, J.-Y. Jung, C. Wang, E. Niforatos, and A. Bozzon. "How Emoji and Explanations Influence Adherence to AI Recommendations". In: *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. IVA '23. Würzburg, Germany: Association for Computing Machinery, 2023. DOI: 10.1145/3570945.3607300

- S. Kernan Freire, C. Wang, S. Ruiz-Arenas, and E. Niforatos. "Tacit Knowledge Elicitation for Shop-Floor Workers with an Intelligent Assistant". In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA '23. Hamburg, Germany: Association for Computing Machinery, 2023. DOI: 10.1145/3544549.3585755

- M. Foosherian, S. Kernan Freire, E. Niforatos, K. A. Hribernik, and K.-D. Thoben. "Break, Repair, Learn, Break Less: Investigating User Preferences for Assignment of Divergent Phrasing Learning Burden in Human-Agent Interaction to Minimize Conversational Breakdowns". In: MUM '22. Lisbon, Portugal: Association for Computing Machinery, 2022, pp. 151–158. DOI: 10.1145/3568444.3568454

- S. Kernan Freire, E. Niforatos, Z. Rusak, D. Aschenbrenner, and A. Bozzon. "A Conversational User Interface for Instructional Maintenance Reports". In: *Proceedings of the 4th Conference on Conversational User Interfaces*. CUI '22. Glasgow, United Kingdom: Association for Computing Machinery, 2022. DOI: 10.1145/3543829.3544516