Harnessing Large Language Models for Cognitive Assistants in Factories

Kernan Freire, Samuel; Foosherian, Mina; Wang, Chaofan; Niforatos, Evangelos

DOI
10.1145/3571884.3604313

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 5th International Conference on Conversational User Interfaces, CUI 2023

Harnessing Large Language Models for Cognitive Assistants in Factories

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

Mina Foosherian
BIBA - Bremer Institut für Produktion und Logistik GmbH
Bremen, Germany
fos@biba.uni-bremen.de

C. Wang
Delft University of Technology
Delft, The Netherlands
c.wang-16@tudelft.nl

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

ABSTRACT

As agile manufacturing expands and workforce mobility increases, the importance of efficient knowledge transfer among factory workers grows. Cognitive Assistants (CAs) with Large Language Models (LLMs), like GPT-3.5, can bridge knowledge gaps and improve worker performance in manufacturing settings. This study investigates the opportunities, risks, and user acceptance of LLM-powered CAs in two factory contexts: textile and detergent production. Several opportunities and risks are identified through a literature review, proof-of-concept implementation, and focus group sessions. Factory representatives raise concerns regarding data security, privacy, and the reliability of LLMs in high-stake environments. By following design guidelines regarding persistent memory, real-time data integration, security, privacy, and ethical concerns, LLM-powered CAs can become valuable assets in manufacturing settings and other industries.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Natural language interfaces**; • **Computing methodologies → Natural language generation**; • **Information systems → Users and interactive retrieval**; • **Applied computing → Industry and manufacturing**.

KEYWORDS

cognitive assistant, conversational user interfaces, knowledge management, industry 5.0, human-centered AI, knowledge sharing

ACM Reference Format:
S. Kernan Freire, Mina Foosherian, C. Wang, and E. Niforatos. 2023. Harnessing Large Language Models for Cognitive Assistants in Factories. In *ACM conference on Conversational User Interfaces (CUI '23), July 19–21, 2023, Eindhoven, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3571884.3604313

CUI '23, July 19–21, 2023, Eindhoven, Netherlands

ACM ISBN 979-8-4007-0014-9/23/07.
https://doi.org/10.1145/3571884.3604313

1 INTRODUCTION

In the era of Industry 5.0 [38], the complexities of agile production lines and a highly mobile workforce have emphasized the need for efficient knowledge sharing among factory workers. Knowledge management has become a critical aspect of modern manufacturing environments, and leveraging technology to facilitate this process is paramount. One emerging solution is using Cognitive Assistants (CAs), which can empower productivity and bridge knowledge gaps by learning from experienced workers and sharing relevant information within the workplace [1, 2, 20].

Enhancing CAs with Large Language Models (LLMs), such as GPT-3.5, offers numerous opportunities to improve their abilities in knowledge sharing and support across various domains. This paper explores the opportunities, barriers, and user acceptance of CAs powered by GPT-3.5 in manufacturing settings. We conducted a literature review, designed, developed, and tested a proof-of-concept for two factory contexts, and collaborated with factory representatives through focus group sessions to gain insights into the perceptions and expectations of LLM-empowered CAs in manufacturing.

2 LITERATURE REVIEW

2.1 Cognitive Assistants and Their Applications

Cognitive assistants, unlike systems designed to replace humans in specific tasks (e.g., industrial robots), aim to augment human capabilities in complex tasks, such as supporting lifelong education and machine operation [1, 2, 20]. These assistants often excel in efficient communication, reliably and consistently interacting with multiple users simultaneously [1, 2]. To achieve these goals, cognitive assistants should facilitate efficient human-machine interaction through natural language processing, gesture interpretation, perception, vision, sound recognition, augmented reality, and other techniques [1, 2].

Among these methods, conversational agent-based natural language communication is the most prevalent interaction technique for cognitive assistants, encompassing natural language understanding, generation, and dialogue [1]. Conversational agents engage users in natural language and can efficiently perform labor-intensive tasks across various industries, such as customer service, healthcare, education, e-banking, and personal assistance [7, 19, 23, 40].

Industrial applications of cognitive assistants, akin to Alexa, Google Assistant, or Siri, are an emerging research area. These prototypes have appeared under various names, such as intelligent (virtual/personal) assistants, digital assistants, software robots, or simply chatbots. Cognitive assistants in manufacturing offer significant advantages, including centralized access to heterogeneous information systems, task delegation, and hands- and gaze-free interactions during work [37]. Moreover, cognitive assistants can be employed for on-the-job worker training [18] and for adjusting machine parameters in simulations [22].

2.2 Leveraging Cognitive Assistants for Knowledge Sharing in Manufacturing Settings

Much of the literature concerning cognitive assistants in manufacturing emphasizes knowledge and information delivery. Examples include context-aware assistance [12], predictive maintenance recommendations [36], and decision support based on shop-floor data analytics [3, 27]. For instance, Rodriguez et al. [29] present a mixed reality assistance system for real-time assembly operations. The system evaluates the operation context using a recognition system that identifies the completion of each assembly step to generate the subsequent instruction. Belkadi et al. [6] propose a context-aware, knowledge-based system to support manufacturing operators by integrating knowledge management, context management, and simulation management modules, aiding real-time worker decision-making. Similar to Rodriguez et al. [29], context management obtains contextual information to comprehend the user's situation and implements simulation techniques to anticipate the effects of worker decisions. Büttner et al. [9] employ a hand-tracking algorithm to detect incorrect picking actions and assembly process errors. Tao et al. [33] utilize wearable and environmental sensing devices to capture worker activity, guiding task execution. Josifovska et al. [16] incorporate a context manager module, including a human digital twin, to simulate specific human abilities and preferences.

The literature has also explored knowledge acquisition from workers, albeit to a significantly lesser degree. However, we believe that it is crucial for cognitive assistants' long-term success on the shop floor. Technological probes have shown that supporting workers to think aloud can result in the creation of high-quality reports from which knowledge can be extracted [17]. Fenoglio et al. [11] suggest a system for capturing explicit and tacit knowledge (through best practices) from experienced industrial workers, employing a role-playing game where a virtual agent interacts with human experts and knowledge engineers to iteratively extract and represent knowledge. Although the system can capture tacit knowledge, it requires human intervention. Fully capturing tacit knowledge (typically nonverbal and unexpressed) with a purely algorithmic approach is deemed impossible by Fenoglio et al. [11]. Similarly, Soliman and Vanharanta [30] propose a knowledge creation and retention model using artificial intelligence, but no practical application has been reported in the literature. Hoerner et al. [13] suggest a digital assistance system to support operator troubleshooting processes on the shop floor, developing a method for capturing and structuring expert tacit knowledge. However, this method is not applied by the digital assistant but by human knowledge managers responsible for extracting, representing, and delivering the acquired knowledge as the system input. Interestingly, state-of-the-art natural language processing (NLP) techniques, such as recent large model Models (LLM), may be powerful enough to (partially) fulfil the role of knowledge manager.

2.3 Large Language Models

Large language models have recently garnered substantial interest in the field of natural language processing and beyond. Scaling up LLMs, for instance, by augmenting model parameters, enhances performance and sample efficiency across diverse NLP tasks [35]. Moreover, bigger LLMs exhibit emergent capabilities absent in smaller counterparts [35]. One notable ability is zero-shot prompting, in which a pre-trained language model can tackle tasks using natural language instructions as prompts without additional training or parameter adjustments [8, 35]. LLMs also demonstrate remarkable few-shot prompting or in-context learning skills, where the model improves performance on a downstream task by conditioning on a prompt containing input-output examples [35]. Leveraging zero-shot/few-shot prompting, companies can craft custom prompts to generate responses tailored to their requirements. Furthermore, several LLMs can integrate other NLP tasks, such as text translation and simplification, to better address user needs.

LLMs' pre-training on large-scale, mixed-source corpora enables them to capture extensive knowledge from the data [39]. As a result, recent research has focused on utilizing LLMs for domain-specific tasks and assessing their adaptability [39]. Various studies have investigated the application of GPT-3.5 and other LLMs in the medical field, encompassing areas such as biological information extraction [32], medical consultation advice [25], and report simplification [14, 39]. Concurrently, multiple empirical studies have shown LLMs to be effective writing or reading assistants in educational contexts [5, 24, 39]. Furthermore, several studies have employed LLMs to address diverse legal tasks, including legal document analysis, judgment prediction, and document writing [34, 39].

However, LLMs may generate text that is semantically plausible and syntactically correct but factually wrong, a phenomenon, known as "hallucination" [4]. The suitability of the term hallucination is questionable as it might imply changes in one's perceptual experience, which LLMs do not have. However, considering its widespread adoption, we will continue to use this term in this paper for clarity. Ji et al. [15] differentiate intrinsic and extrinsic hallucinations in natural language generation. The former term refers to contradictions between source material (e.g., training data and prompts), whereas extrinsic hallucinations refer to information that cannot be verified by the source material. Whereas intrinsic hallucinations are clearly erroneous, extrinsic hallucinations can still be an issue in high-stake contexts where all outputs should be verifiable.

2.4 Exploring Large Language Models for Domain-Specific Knowledge Sharing

In order to augment an LLM with domain-specific or custom knowledge, two main strategies exist. The first method, fine-tuning, involves training the model further using a custom data set [28].

However, such models might struggle to effectively use the context surrounding entities, resulting in limitations when fine-tuning on smaller domain-specific data sets [10]. Moreover, while fine-tuning can achieve high accuracy and comprehensiveness, it requires significant time and resources for both training and hosting the custom model. Alternatively, in-context learning offers relevant information from the custom data set in connection with the user query during the query process, thus improving the model's performance [21]. Lewis et al. [21] presented the retrieval-augmented generation (RAG) architecture, which operates similarly to a conventional seq2seq model by taking a single sequence as input and producing a corresponding output sequence. RAG enhances its performance by using the input to retrieve a set of pertinent documents from Wikipedia rather than directly passing the input directly to the generator.

In addition to retrieving information from domain-specific or custom databases, LLM-driven systems can create implicit data to enhance these databases further. Such a system powered by an LLM could allow users to select question-and-answer pairs for inclusion or to submit questions that the assistants are unable to answer to these databases [31]. Over time, LLM-driven systems will amass a considerable collection of answers and become increasingly adept as primary knowledge specialists.

To the best of our knowledge, no cognitive assistant currently exists that leverages an LLM and is specifically designed to tackle inherent weaknesses of LLMs, such as hallucinations, by learning from skilled workers and disseminating pertinent knowledge in manufacturing settings. Consequently, we aim to investigate the possibilities, obstacles, and user receptiveness of cognitive assistants employing LLMs in the context of manufacturing.

3 SYSTEM DESCRIPTION

To explore the potential of using LLMs to augment cognitive assistants in manufacturing, we built LLM-powered features for existing cognitive assistants. The assistants originate from two different manufacturing contexts, namely textile and detergent production, communicating in Italian and Dutch, respectively. Both assistants can be accessed in a smartphone application and feature a conversational user interface with speech and text capabilities for flexible user interaction. In this section, we describe the architecture, components, and user interaction related to the new LLM-powered features.

3.1 Manufacturing context

The assistant for the textile production context is designed to support novice workers with basic tasks, for example, by answering FAQs about how to operate the looms or perform tasks such as knotting. As such, the knowledge base upon which the assistant relies is relatively static. In contrast, the assistant for the detergent production context aims to share more complex knowledge between the workers, for example, how to solve emergent issues with the production line. This type of knowledge is more likely to change over time as it is affected by raw ingredient quality, machine wear, and product characteristics. As such, it is important to continuously incorporate newly discovered knowledge. The LLM-powered features replace existing capabilities in the aforementioned assistants. Specifically, capabilities that originally provided issue support derived from past issue reports, and answering frequently asked questions.

3.2 Architecture and user interaction

The LLM-powered features are used to construct a multi-part prompt which is sent to the GPT-3.5 API[1]. The prompts consist of a predefined system instruction, existing domain-specific texts, the user's request, and query results from a knowledge graph (only for the detergent production assistant, see Figure 1).

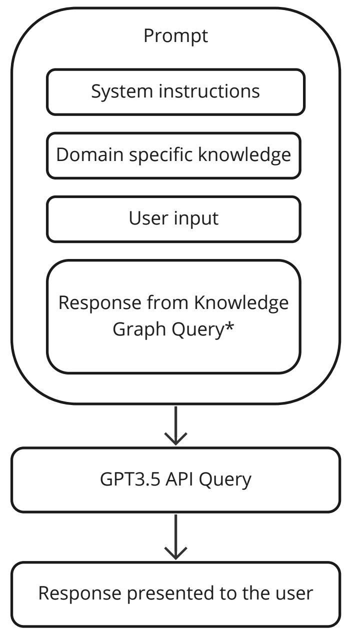

Figure 1: Prompt Engineering

The 150-word system instruction contains statements such as "*You are a helpful assistant for factory workers*", "*you provide concise answers in a maximum of 50 words*", a general description of the involved machines and process. This is followed by the factory and domain-specific information. In the case of the detergent factory, this is 2,000 words of previously collected issue descriptions, accounting for about 30 different issues. For the textile production assistant, we included the text from seven learning modules and ten FAQ pairs. Then, we include the user's input. For the detergent production assistant, the user is asked to provide a description of the symptoms they observed, including the main machine component involved and the error code. This information is used to query a knowledge graph for any previously recorded issue solutions. Any result is appended to the prompt. For the textile production assistant, we simply include the user's question since a knowledge graph is not available.

[1]https://platform.openai.com/docs/guides/chat/introduction—last accessed June 8, 2023

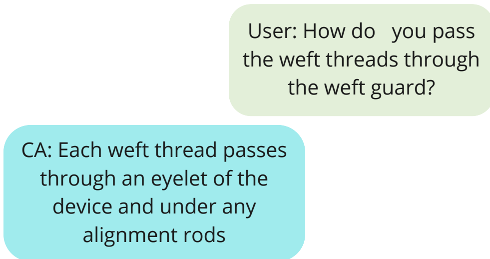

Figure 2: A dialogue showing a user asking for task instructions from the cognitive assistant (CA)

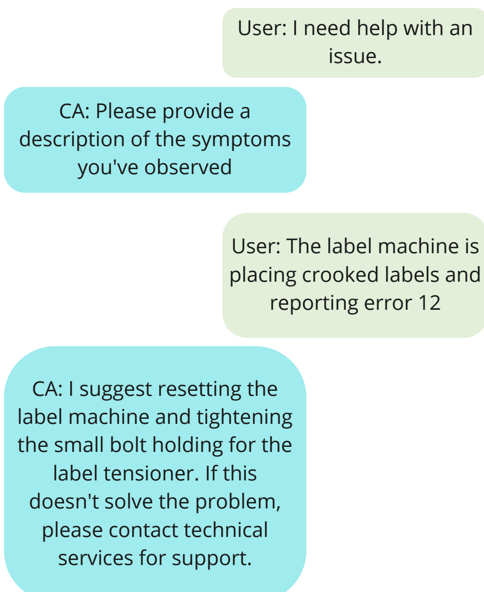

Figure 3: A dialogue showing a user requesting help with a machine issue from the cognitive assistant (CA)

4 EVALUATION METHOD

4.1 User Study with Factory Representatives

During the evaluation's initial phase, we assessed the accuracy of the LLM-powered features' responses. We enlisted three representatives from a detergent factory and two from a textile factory. In individual sessions, participants posed various questions related to their production lines using the LLM-powered features and evaluated the accuracy of the responses. The participants were free to pose any questions they deemed fit, for example, how to solve a specific issue at the production line or how to conduct a standard operating procedure. All participants had prior experience using non-LLM-powered cognitive assistants.

4.2 Feature Demonstration and Focus Group

In the evaluation's second phase, we showcased the LLM-powered features to 22 individuals, comprising representatives from three different factories, developers, and researchers, including three participants from the first phase. Subsequently, we explored the potential advantages, drawbacks, and constraints of employing LLM-powered assistants in the targeted factories and the broader manufacturing sector. All participants were familiar with cognitive assistants and the manufacturing context.

5 RESULTS AND DISCUSSION

Overall, the factory representatives were impressed by the accuracy of the responses. Furthermore, they were satisfied with the robustness of the system's ability to interpret some of their (poorly phrased) questions, a common issue with non-LLM-powered cognitive assistants. However, they noted that they sometimes received a generic response that was not useful. This usually occurred when there was insufficient information provided in the prompt for that particular topic.

The factory representatives expressed several concerns related to the use of such technology in manufacturing environments. One significant apprehension is the possibility of confidential information and proprietary knowledge being exposed to competitors through the LLM's API and training corpus. This risk can jeopardize the competitive advantage of manufacturing companies and compromise intellectual property. Furthermore, they did not want to have their company's name or IP address associated with all prompts made by their employees, especially if it may somehow negatively impact their company's brand image.

However, it should be noted that most of the participants initially believed that all the data sent to the API was being used by OpenAI to improve their model. After all, this is the case for the consumer-facing, browser-based interface commonly known as chatGPT, albeit with an opt-out option[2]. Conversely, the API follows an opt-in protocol and does not use prompts for training purposes by default[3]. Despite knowing this, many of the factory representatives were cautious of trusting another corporation with their data. As such, they were more welcoming to using open-source, locally-hosted LLMs. These solutions could offer more control and privacy while mitigating potential risks and exposure to competitors. Although open-source LLMs are available, such as LLaMa and its derivatives such as Alpaca, they pale in comparison to the abilities of GPT-3.5 and GPT-4. However, some experimentation with ChatGLM appears to be more promising.

One researcher pinpointed issues such as (artificial) hallucinations, elementary reasoning errors, privacy invasion, guarantees on recommendations or correctness of explanations, and the danger of employing LLMs where lives or money are on the line. For example, in the detergent factory where volatile chemicals are mixed. These concerns further highlight the need for cautious and responsible implementation of LLM-powered CAs in manufacturing settings to keep potential risks and drawbacks in check. These points echo the

[2] https://help.openai.com/en/articles/7039943-data-usage-for-consumer-services-faq—last accessed June 8, 2023
[3] https://openai.com/policies/api-data-usage-policies—last accessed June 8, 2023

main limitations that OpenAI [26] have outlined in their technical report:

(1) does not learn from experience
(2) not fully reliable
(3) limited context window
(4) risks: bias, misinformation, over-reliance, privacy, cyber security, and more

All of these limitations are (partially) mitigated by providing the assistant with persistent memory, for example, in the form of a knowledge graph as we demonstrated for the detergent business case. This enables the assistant to access knowledge specifically for its context of use (e.g., a factory), learn from its users, and maximize the potential of its context window by querying knowledge relevant to the situation at hand. Ultimately, this will also improve the reliability of its responses. Furthermore, if the user suspects that the LLM is hallucinating, the cognitive assistant could bypass it and retrieve the raw data from the knowledge base for the user to check.

From the perspective of the designers and developers of conversational systems, using LLMs can greatly reduce the resources needed for the development and maintenance of assistants as their features do not have to be programmed in as much detail as it was previously necessary (e.g., intent-based systems).

6 DESIGN GUIDELINES

To maximize the effectiveness of LLM-powered CAs, the following design guidelines should be considered:

- **Integrate dynamic domain knowledge**. Integrate persistent memory of domain knowledge for the assistant, for example, in the form of a continuously expanding knowledge base. Once the knowledge base becomes too large to fit in one query, techniques such as semantic search can be used to select relevant documents. Documents can be indexed in a vector form to facilitate semantic search. Additionally, the outcomes from prior queries can be saved to reduce both the computational and time expense by repeatedly prompting the large language model (LLM) for the same question.
- **Supply real-time data**. Seamless integration with existing factory systems and machine data can help the LLM make contextual connections between the user's question, the current situation, and knowledge stored in its persistent memory (e.g., in a knowledge graph). Furthermore, this would enable it to calculate production statistics and visualize insights into graphs.
- **Provide transparency** where possible and when requested. Although the output of the LLM cannot (yet) be explained, it is possible to report and explain the (hidden) contents of the prompts. For example, responses from a knowledge base that are included in the prompt could be exposed to the user to improve explainability.
- **Design for security and privacy**. Addressing concerns about data security and privacy is critical for user acceptance and trust in such systems. Solutions such as locally hosted LLMs can help alleviate these concerns.
- **Ethics** Adopting LLM-powered cognitive assistants in manufacturing settings raises ethical concerns, including the potential for hallucinations, lack of reliable explanations, and loss of intellectual property. These can have negative consequences when applied in high-stake business environments. Therefore, it is imperative to maintain transparency regarding potential risks and limitations associated with LLM-powered CAs to mitigate these issues, allowing end-users to make informed decisions on their use.
- **Collect feedback and evaluate**. Regular evaluation and user feedback should be incorporated into the system development process to continuously refine and improve the LLM-powered feature of the cognitive assistant.

7 CONCLUSION

LLM-powered CAs hold promise for improving knowledge sharing and supporting various aspects of manufacturing processes. By addressing challenges related to technical limitations, and ethical concerns, for example, by enabling persistent memory, LLM-powered CAs can become valuable assets in modern manufacturing environments and other industries. Nevertheless, considering the substantial potential of LLMs and the ongoing discovery of their extensive impact, both favorable and unfavorable, it is imperative to engage in critical reflection prior to deployment. The aforementioned design guidelines can serve as "reflection prompts" to facilitate this process.

ACKNOWLEDGMENTS

This work was supported by the European Union's Horizon 2020 research and innovation program via the project COALA "COgnitive Assisted agile manufacturing for a LAbor force supported by trustworthy Artificial Intelligence" (Grant agreement 957296).

REFERENCES

[1] 2016. *Intelligent Cognitive Assistants: Workshop Summary and Recommendations.* https://www.nsf.gov/crssprgm/nano/reports/2016-1003_ICA_Workshop_Final_Report_2016.pdf
[2] 2018. *Intelligent Cognitive Assistants: Workshop Summary and Research Needs.* https://www.nsf.gov/crssprgm/nano/reports/ICA2_Workshop_Report_2018.pdf
[3] Brunno Abner, Ricardo J. Rabelo, Saulo P. Zambiasi, and David Romero. 2020. Production Management as-a-Service: A Softbot Approach. In *Advances in Production Management Systems. Towards Smart and Digital Manufacturing*, Bojan Lalic, Vidosav Majstorovic, Ugljesa Marjanovic, Gregor von Cieminski, and David Romero (Eds.). IFIP Advances in Information and Communication Technology, Vol. 592. Springer International Publishing, Cham, 19–30. https://doi.org/10.1007/978-3-030-57997-5{_}3
[4] Hussam Alkaissi and Samy I McFarlane. 2023. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus* 15, 2 (2023).
[5] Zeljana Basic, Ana Banovac, Ivana Kruzic, and Ivan Jerkovic. 2023. Better by you, better than me, chatgpt3 as writing assistance in students essays. arXiv:2302.04536 [cs.AI]
[6] Farouk Belkadi, Mohamed Anis Dhuieb, José Vicente Aguado, Florent Laroche, Alain Bernard, and Francisco Chinesta. 2020. Intelligent assistant system as a context-aware decision-making support for the workers of the future. *Computers & Industrial Engineering* 139 (Jan. 2020), 105732. https://doi.org/10.1016/j.cie.2019.02.046
[7] Luka Bradeško, Michael Witbrock, Janez Starc, Zala Herga, Marko Grobelnik, and Dunja Mladenić. 2017. Curious Cat–Mobile, Context-Aware Conversational Crowdsourcing Knowledge Acquisition. *ACM Trans. Inf. Syst.* 35, 4, Article 33 (aug 2017), 46 pages. https://doi.org/10.1145/3086686
[8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
[9] Sebastian Büttner, Oliver Sand, and Carsten Röcker. 2017. Exploring Design Opportunities for Intelligent Worker Assistance: A New Approach Using

Projetion-Based AR and a Novel Hand-Tracking Algorithm. In *Ambient Intelligence (Lecture Notes in Computer Science)*, Andreas Braun, Reiner Wichert, and Antonio Maña (Eds.). Springer International Publishing, Cham, 33–45. https://doi.org/10.1007/978-3-319-56997-0_3

[10] Keyur Faldu, Amit Sheth, Prashant Kikani, and Hemang Akbari. 2021. KI-BERT: Infusing knowledge context for better language and domain understanding. *arXiv preprint arXiv:2104.08145* (2021).

[11] Enzo Fenoglio, Emre Kazim, Hugo Latapie, and Adriano Koshiyama. 2022. Tacit knowledge elicitation process for industry 4.0. *Discover Artificial Intelligence* 2, 1 (March 2022), 6. https://doi.org/10.1007/s44163-022-00020-w

[12] Markus Funk, Tilman Dingler, Jennifer Cooper, and Albrecht Schmidt. 2015. 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* (Osaka, Japan) *(UbiComp/ISWC'15 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1269–1273. https://doi.org/10.1145/2800835.2807942

[13] Lorenz Hoerner, Markus Schamberger, and Freimut Bodendorf. 2022. Using Tacit Expert Knowledge to Support Shop-floor Operators Through a Knowledge-based Assistance System. *Computer Supported Cooperative Work (CSCW)* (Sept. 2022). https://doi.org/10.1007/s10606-022-09445-4

[14] Katharina Jeblick, Balthasar Schachtner, Jakob Dexl, Andreas Mittermeier, Anna Theresa Stüber, Johanna Topalis, Tobias Weber, Philipp Wesp, Bastian Sabel, Jens Ricke, and Michael Ingrisch. 2022. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. arXiv:2212.14882 [cs.CL]

[15] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* 55, 12, Article 248 (mar 2023), 38 pages. https://doi.org/10.1145/3571730

[16] Klementina Josifovska, Enes Yigitbas, and Gregor Engels. 2019. A Digital Twin-Based Multi-modal UI Adaptation Framework for Assistance Systems in Industry 4.0. In *Human-Computer Interaction. Design Practice in Contemporary Societies (Lecture Notes in Computer Science)*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 398–409. https://doi.org/10.1007/978-3-030-22636-7_30

[17] Samuel Kernan Freire, Evangelos Niforatos, Zoltan Rusak, Doris Aschenbrenner, and Alessandro Bozzon. 2022. A Conversational User Interface for Instructional Maintenance Reports. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 5, 6 pages. https://doi.org/10.1145/3543829.3544516

[18] Samuel Kernan Freire, Sarath Surendranadha Panicker, Santiago Ruiz-Arenas, Zoltán Rusák, and Evangelos Niforatos. 2022. A Cognitive Assistant for Operators: AI-Powered Knowledge Sharing on Complex Systems. *IEEE Pervasive Computing* (2022), 1–9. https://doi.org/10.1109/MPRV.2022.3218600

[19] Everlyne Kimani, Kael Rowan, Daniel McDuff, Mary Czerwinski, and Gloria Mark. 2019. A conversational agent in support of productivity and wellbeing at work. In *2019 8th international conference on affective computing and intelligent interaction (ACII)*. IEEE, 1–7.

[20] Nguyen-Thinh Le and Laura Wartschinski. 2018. A cognitive assistant for improving human reasoning skills. *International Journal of Human-Computer Studies* 117 (2018), 45–54.

[21] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401 [cs.CL]

[22] Franz Georg Listl, Jan Fischer, and Michael Weyrich. 2021. 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, 1–4. https://doi.org/10.1109/ETFA45728.2021.9613681

[23] Bei Luo, Raymond YK Lau, Chunping Li, and Yain-Whar Si. 2022. A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 12, 1 (2022), e1434.

[24] Kamil Malinka, Martin Perešíni, Anton Firc, Ondřej Hujňák, and Filip Januš. 2023. On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree? arXiv:2303.11146 [cs.CY]

[25] Oded Nov, Nina Singh, and Devin Mann. 2023. Putting ChatGPT's Medical Advice to the (Turing) Test. arXiv:2301.10035 [cs.HC]

[26] OpenAI. 2023. *GPT-4 Technical Report*. Technical Report. OpenAI. https://cdn.openai.com/papers/gpt-4.pdf

[27] Ricardo J. Rabelo, Saulo Popov Zambiasi, and David Romero. 2019. Collaborative Softbots: Enhancing Operational Excellence in Systems of Cyber-Physical Systems. In *Collaborative Networks and Digital Transformation*, Luis M. Camarinha-Matos, Hamideh Afsarmanesh, and Dario Antonelli (Eds.). IFIP Advances in Information and Communication Technology, Vol. 568. SPRINGER NATURE, 55–68. https://doi.org/10.1007/978-3-030-28464-0{_}6

[28] Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. 2019. Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification. *arXiv preprint arXiv:1908.11860* (2019).

[29] Leonardo Rodriguez, Fabian Quint, Dominic Gorecky, David Romero, and Héctor R. Siller. 2015. Developing a Mixed Reality Assistance System Based on Projection Mapping Technology for Manual Operations at Assembly Workstations. *Procedia Computer Science* 75 (Jan. 2015), 327–333. https://doi.org/10.1016/j.procs.2015.12.254

[30] Yehya Soliman and Hannu Vanharanta. 2020. A Model for Capturing Tacit Knowledge in Enterprises. In *Advances in Human Factors, Business Management and Leadership (Advances in Intelligent Systems and Computing)*, Jussi Ilari Kantola and Salman Nazir (Eds.). Springer International Publishing, Cham, 141–148. https://doi.org/10.1007/978-3-030-20154-8_14

[31] Quy Tang. 2023. *Integrating ChatGPT with internal knowledge base and question-answer platform*. https://medium.com/singapore-gds/integrating-chatgpt-with-internal-knowledge-base-and-question-answer-platform-36a3283d6334

[32] Ruixiang Tang, Xiaotian Han, Xiaoqian Jiang, and Xia Hu. 2023. Does Synthetic Data Generation of LLMs Help Clinical Text Mining? arXiv:2303.04360 [cs.CL]

[33] Wenjin Tao, Ze-Hao Lai, Ming C. Leu, Zhaozheng Yin, and Ruwen Qin. 2019. A self-aware and active-guiding training & assistant system for worker-centered intelligent manufacturing. *Manufacturing Letters* 21 (Aug. 2019), 45–49. https://doi.org/10.1016/j.mfglet.2019.08.003

[34] Dietrich Trautmann, Alina Petrova, and Frank Schilder. 2022. Legal Prompt Engineering for Multilingual Legal Judgement Prediction. arXiv:2212.02199 [cs.CL]

[35] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682* (2022).

[36] Stefan Wellsandt, Mina Foosherian, Katerina Lepenioti, Mattheos Fikardos, Gregoris Mentzas, and Klaus-Dieter Thoben. 2022. Supporting Data Analytics in Manufacturing with a Digital Assistant. In *Advances in Production Management Systems. Smart Manufacturing and Logistics Systems: Turning Ideas into Action*, Duck Young Kim, Gregor von Cieminski, and David Romero (Eds.). IFIP Advances in Information and Communication Technology, Vol. 664. Springer Nature Switzerland, Cham, 511–518. https://doi.org/10.1007/978-3-031-16411-8{_}59

[37] Stefan Wellsandt, Karl Hribernik, and Klaus-Dieter Thoben. 2021. Anatomy of a Digital Assistant. In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*, Alexandre Dolgui, Alain Bernard, David Lemoine, Gregor von Cieminski, and David Romero (Eds.). IFIP Advances in Information and Communication Technology, Vol. 633. Springer International Publishing, Cham, 321–330. https://doi.org/10.1007/978-3-030-85910-7{_}34

[38] Xun Xu, Yuqian Lu, Birgit Vogel-Heuser, and Lihui Wang. 2021. Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems* 61 (2021), 530–535. https://doi.org/10.1016/j.jmsy.2021.10.006

[39] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. arXiv:2303.18223 [cs.CL]

[40] Shuo Zhou and Timothy Bickmore. 2022. A Virtual Counselor for Breast Cancer Genetic Counseling: Adaptive Pedagogy Leads to Greater Knowledge Gain *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 436, 17 pages. https://doi.org/10.1145/3491102.3517553