ESAIM 2024 – 2nd European Symposium on Artificial Intelligence in Manufacturing

CPPM Copilot: Proposing an AI-based assistant for manual assembly tasks in a flexible production

Jonathan Nussbaum[1,*], Tatjana Legler[1,2], and Martin Ruskowski[1,2]

[1] Chair of machine tools and control systems, University of Kaiserslautern-Landau (RPTU), 67663 Kaiserslautern, Germany
[2] Innovative Factory Systems, German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany

[*] Corresponding author. Tel.: +49-631-205-4256.
E-mail: jonathan.nussbaum@rptu.de

Abstract. With an increase in flexibility in industries adopting the tenets of shared, skill-based, and modular production, a higher variability of products a single manufacturer offers can be anticipated. An immediate consequence is a potential inability to instruct assembly personnel in an adequate and detailed manner. This problem would primarily affect small and medium enterprises engaged in individual and small series productions. To face the issue, this paper examines the suitability of facilitating an AI assistant to support workers handling an ever-expanding range of assembly tasks. This assistant would be realized through a retrieval-augmented generation system, founded on an Large Language Model (LLM) and a knowledge base. We especially propose a locally trained and hosted LLM, aiming to enhance effective flexibility and applicability while minimizing individual installation and setup times. By strictly formalizing descriptions of assembly or disassembly steps in a knowledge base, manufacturing difficulties can be presented as informational problems, at which LLMs excel. Through this, we are trying to extend the increased efficiency of knowledge workers empowered by utilization of LLMs such as GPT-4 into manufacturing. Using verbal inputs as well as reading the generated feedback back to a worker, we aim to keep a worker engaged with their primary tasks and, furthermore, reduce idle times caused by knowledge gaps. By taking this verbal/auditive-only approach, we secondarily aim towards increasing worker autonomy by answering miscellaneous workplace-related questions alongside knowledge-based problems of the day-to-day business.

Keywords: Worker assistance · Large Language Models · AI in manufacturing · Capability-Skill-Service model

1 Introduction and Motivation

Humans still hold one of the most important positions inside modern manufacturing shop floor environments, mainly due to their adaptability, their versatility as well as their problem-solving skills. While appearances like *Figures* Figure01 [15] or the new iteration of *Boston Dynamics'* Atlas [4] are aiming to replicate the general dexterity of humans, neither robot is supposed to replace a human in the complex and flexible sectors that are small series or even individual part production, with Figure aiming to support the warehouse and retail workforce and Boston Dynamics being factored in for automotive serial production [7,5]. Neither these two humanoid robots, nor conventional stationary robots, can replace the skilled personnel required to operate agile production lines [8]. Still, these skilled workers face problems with the ever-changing product palette of the modern, flexible shop floors.

Within this paper, we are concerned with just the small-scale problem that is manual product assembly, within which we are focusing on issues regarding assembly instructions. Key challenges regarding manual assembly have been identified by [9] as

- instructions being of poor quality and too general in their nature,
- a detailed, digital documentation being unavailable to staff facing problems,
- and conventional paper-based instructions lacking traceability.

This is by no means an argument for extraordinarily formalized assembly instructions, as a recent study has found human workers preferring to pick a cognitively demanding task, while allocating manual labor to a cobot [12].

Regardless of whether there is a cobot to take up other, repetitive tasks, the mentally demanding tasks of figuring out new and not yet learned assembly instructions is an adequate example for a relevant problem-solving skill. But while a human might more or less enjoy the time spent working out how to assemble a new product, they might benefit from an assistant supporting them. While not solving the workers' problems for them, a smart personal assistant can still be of benefit by increasing problem-solving skills in the long term, as shown by [14]. If applicable, this would enable workers to more easily comprehend instructions initially found to be rather incomprehensible. On the other hand, if there is a possibility to formalize and standardize assembly instructions to a point, where the manual assembly itself is solely considerable to be a problem of knowing the next steps, then the findings of [2] regarding the increase in efficiency of knowledge workers utilizing Generative Artificial Intelligence (GenAI) can be transferred into manual assembly.

2 Context and Background

Following, we highlight the background, concepts, and technologies necessary for our concept. Next to the benefits of verbally communicating assistants, we cover the Capability-Skill-Service model (CSS model) and GenAI as well as Natural Language Processing (NLP).

2.1 Benefits of a verbally communicating AI assistant

In a study, [13] have found difficulties when integrating a worker assistance system relying on visual input for the personnel. Workers with cognitive disabilities sometimes forgot to use their assistance systems. Experienced workers even chose to disregard system instructions, while still appreciating the systems' potential for alerting them in cases of potential assembly errors. Furthermore, their study showed problems when using, for example, a Microsoft HoloLens or projectors for visual assistance, stating the HoloLens' limited battery life and a projectors' increased need for maintenance.

Considering an approach contrary to using visual interactions between personnel and an assistance system, a system could be interactive with on a voice-first basis. A voice-first approach would – in the case of worker assistance – mean using verbal outputs of a human as the primary system input. Advantages of such a voice controlled system are the flexibility to individually adapt them to any environment or worker, as well as being user-agnostic and therefore open to anybody willing to use it. [11]

Taking just these two positions into account and, furthermore, [11]'s statement of AI-based personal assistants having a high potential when a user would need to interface with machinery as well as [2]'s values regarding the increase in performance observed in knowledge workers when utilizing an AI assistant (12.2 % more tasks completed, being 25.1 % faster, while achieving 40 % better results), leads to the premise of an AI-based and voice controlled assistance system. Notably, the averaged values reported by [2] were drastically lower for already better-than-average workers, while still being a net positive, but – more importantly – the increase in performance was even higher for workers below the average initial performance levels.

2.2 Capability-Skill-Service model

The CSS model is a conceptual framework that unifies the terminology and defines a vocabulary for capabilities, skills, and services in the context of production processes. A simplified overview of the model is shown in fig. 1. It aims to enhance understanding and interoperability in new production concepts and support standardization activities in the manufacturing industry. The model is an extension of the Product-Process-Resource representation paradigm and focuses on capturing functions at different levels: Capabilities represent functions in production process steps; skills are implementations of these functions provided by specific resources; and services define offerings of capabilities in broader supply chain networks.

In the scope of the concept proposed in this work, the relevant parts of fig. 1 are the offered **service**, which provides the offered **capability** and the **skill** realizing it, as well as the **resource** providing said capability and skill. From a top-down view, the service offered is an outside representation and offer towards other agents in need of the specific service. A capability, provided by the service and the resource, is a representation of an action, complete with options

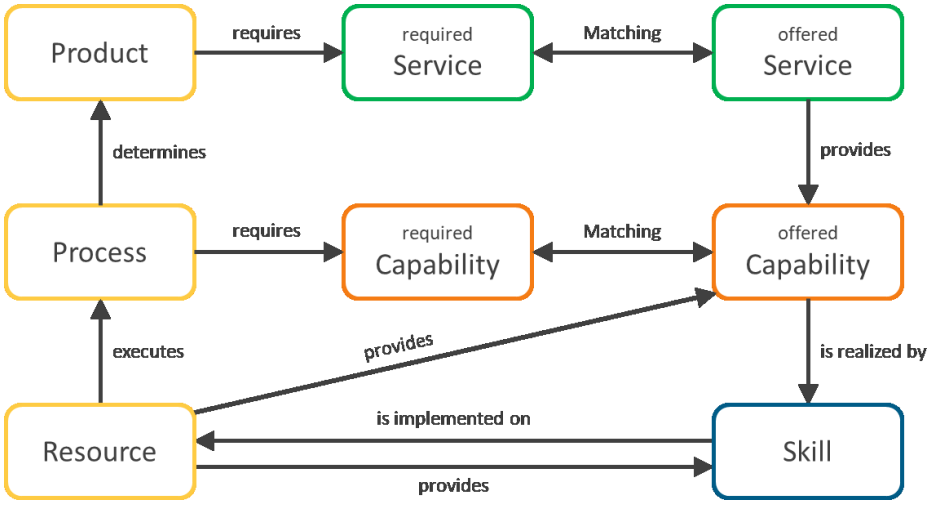

Fig. 1. Simplified overview on most important aspects of the CSS model [3].

for parametrization and beforehand information regarding, example given, runtimes or, ever-more important, energy and material consumption or greenhouse gas emissions. As a capability is an abstract entity provided by said services and provided by certain resources but not tied to either definitely, if there were multiple resources providing a certain capability, the service would be the provider for each of them. The implementation of a capability down into the machine and module level is the so-called skill. [3]

2.3 Generative Artificial Intelligence and Natural Language Processing

Generative Artificial Intelligence is a term used colloquially for Artificial Intelligence (AI) capable of generating new information. Common tasks for GenAI stem from the field of NLP, taking the form of text interpretation and generation. Other media generated are songs, images, as well as videos. The generation follows a user input, which is often given in text form, optionally supplemented by documents, web pages, images or sound files. The topics covered by this diverge greatly, with a user being able to extract information about topics ranging from culinary advice, over an overview of Albrecht Dürer's Fechtbuch, through an explanation of general relativity fit for elementary school kids, to details regarding internationally less known tabletop role-playing games. [6]

NLP is the field of automatic analysis and representation of human language. It is concerned with understanding, interpreting and generating human language on a meaningful scale. Further NLP tasks are, for example, speech recognition, translation, or summarization. [10]

Using AI networks for NLP tasks has been a long-going trend, with statistical language models and then neural language models being having been used for

NLP tasks. Further development resulted in pre-trained language models trained on extensive datasets, the largest of which are known as **Large Language Models (LLMs)**. The emerging abilities of the models were: a sharp increase in general performance; suddenly being able to generate very high-quality text samples; possessing robust learning; and reasoning abilities.

3 Proposed concept

The generalized description of the concept is summarized in fig. 2. A workers' spoken query is first picked up by a headset connected to the module and, using NLP, transformed into text. Semantic information is then extracted from this text by an LLM, before it is matched with information from the knowledge base, possibly utilizing knowledge graphs. From the knowledge base, information is retrieved and through the LLM expressed in natural text. This text is then turned into speech using NLP and played back to the worker. The separation of NLP and LLMs, as suggested in fig. 2 is not strictly necessary, as LLMs tend to be generally efficient in NLP tasks. A separation might still prove useful, in case information modelling is a task utilized best by an LLM without additional tasks.

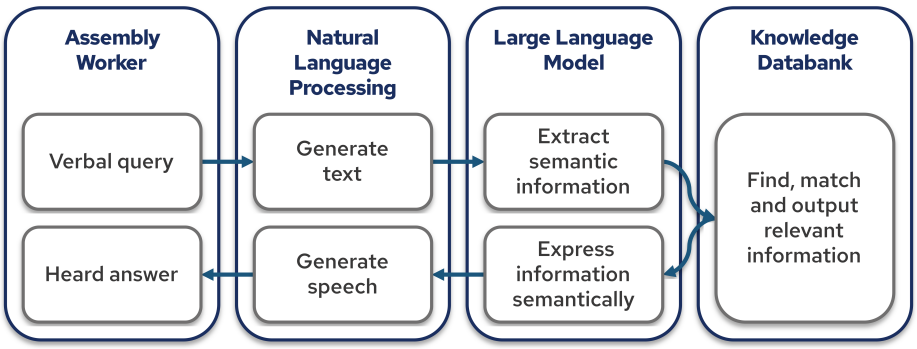

Fig. 2. Overview of the core concept.

3.1 Test demonstrator and demo use-case

During a previous project, 'KoKoBot – Setup of a collaborative and cooperative robotics platform', a demonstrator consisting of three Cyber-Physical Production Moduless (CPPMs) has been set up. The demonstrator and other project results can be seen in [1]. As one project requirement has been equipping the demonstrator with industrial-grade hardware, it will suit well as an assembly station, noting that one of the three modules is, in fact, called 'manual workstation'. An NVIDIA Jetson Orin already built into the demonstrator is meeting

estimated computational requirements and will therefore be the hardware platform hosting the LLM and NLP applications. Additionally, in case the Jetson Orin proves to be a computational bottleneck, a workstation utilizing an RTX graphics card is part of the module as well, serving as an ample backup. Further equipment needed for the proof of concept would be a product to be used for the assembly or disassembly task. For this, the toy trucks used in the previous project would suffice. After a first proof of concept, more sophisticated assembly groups would undoubtedly be needed. Furthermore, any combination of microphone and speaker would suffice to capture vocal instructions and play back the assistant's replies.

3.2 Foreseeable issues

When integrating an LLM a number of issues have to be addresses upfront. One of these is limiting the system's access to work-unrelated information. While the working personnel should be free to ask anything with a reasonable connection to their work, misconduct should preemptively be inhibited. Furthermore, hallucinations have to be addressed. For this, the proposed concept will initially be used as a test bed, closely monitoring the AI's behavior and enabling timely implementation of preventive measures.

3.3 Integration of the CSS model

When integrating the CSS model with the above concept, the granularity of representing functions becomes a core question. One way of implementation would be to make the complete worker assistance into a single skill. In this case, an additional external input would cause the start of the event chain, beginning with the workers' verbal input. The skill would have no parametrization options, as well as not having any output except for the verbalized response of the system.a

Another, preferable, possibility would be to encapsulate most functionalities into singular skills, calling each other when in execution. This would enable the individual testing of each step of the general concept. Therefore, we propose splitting the functions into the following skills:

Skill 1: A skill that is always running and waiting for a command to start the next Skill.
Skill 2: The skill responsible for NLP in both directions, being configurable to either transcribe spoken text or verbalize textual input, either passing it on to skill three or playing it back to the human.
Skill 3: A skill taking text inputs and using them to fetch requested information from a database, while also capable of translating the database entries into human-readable text, which it transmits back to skill two.

Correct system behavior can thus be tested by using spoken as well as textual standard phrases with skill two, testing the correct textualization and verbalization, and calling the third skill either with a search string or a given database

entry. The worker states a query, which is checked against the manual and a database, and results in a response by the module. Additionally, both skills can be used in other ways, with skill two also able to be used to give the user a heads-up in case of system-wide errors or warnings and skill three being available to different users for questioning the database.

Fig. 3 shows a simplified version of this concept in action. Assuming skill one to be running and triggered by the command word, skill two is used to pick up the workers' query and call skill three. The relevant manual is identified and checked, as well as a database being consulted, before the information is sent back into skill two to be read out to the human.

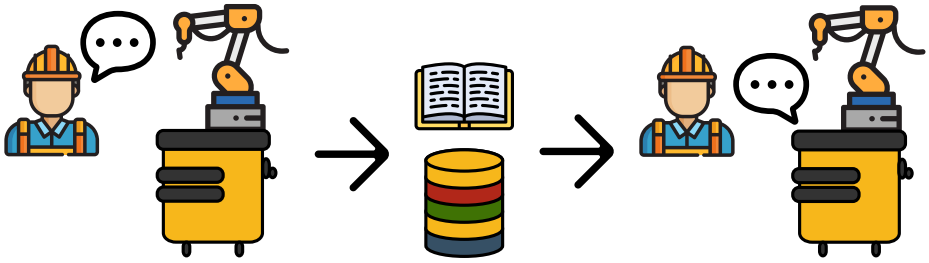

Fig. 3. Simplified depiction of a worker stating a question, a manual and a database being consulted, and the system giving an answer.

4 Conclusion and future directions

Summarized, our concept is aiming to take an assembly workers' verbal input regarding a new product, the assembling steps of which are at least partially unknown. This verbal input is then compiled into a search query for a database containing information regarding the new part. The output of the database search is then compiled into natural language and read back to the human.

This endeavor is only a first step, with multiple possible next steps already figured out. Instead of taking the verbal input for a search query, it might also be a control sequence for the CPPM, triggering different functions and enabling more intuitive control of certain aspects of production. Moreover, the search queries don't necessarily have to relate to the part currently worked on, different other aspects of work might be included in the database, from a guide on how to correctly fill out applications for vacation time to information regarding the current menu in the canteen, stimulating a worker during otherwise possibly boring work and sating the need for certain information. Another aspect of the system is the relatability to different production environments. Human-Robot-Cooperation might be more easily coordinated, if the human can tell the robot what they will do next, as well as giving instructions on what the robot might be supposed to do. The system would likewise suit chemical or pharmaceutical

laboratory environments processing a doctor's or pharmacists recipes as well as more classical shop floors, where part specifications could be checked on the fly without having to leave the working equipment to check the specification.

Acknowledgment

This work was funded by the Carl Zeiss Stiftung, Germany under the Sustainable Embedded AI project (P2021-02-009).

References

1. Project results - department of mechanical and process engineering at rptu (13052024), https://mv.rptu.de/en/fgs/wskl/forschung/kokobot/project-results
2. Dell'Acqua, F., et al.: Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality (2023). https://doi.org/10.2139/ssrn.4573321
3. Diedrich, C., et al.: Information Model for Capabilities, Skills & Services: Definition of terminology and proposal for a technology-independent information model for capabilities and skills in flexible manufacturing. Plattform Industrie 4.0 (2022)
4. Dynamics, B.: Atlas | boston dynamics (4/17/2024), https://bostondynamics.com/atlas/
5. Dynamics, B.: An electric new era for atlas | boston dynamics (4/17/2024), https://bostondynamics.com/blog/electric-new-era-for-atlas/
6. Epstein, Z., et al.: Art and the science of generative ai. https://doi.org/10.1126/science.adh4451
7. FigureAI: About us | figure (4/19/2024), https://www.figure.ai/about-us
8. Freire, S.K., et al.: A cognitive assistant for operators: Ai-powered knowledge sharing on complex systems. IEEE Pervasive Computing **22**(1), 50–58 (2023). https://doi.org/10.1109/MPRV.2022.3218600
9. Johansson, P.E.C., et al.: Challenges of handling assembly information in global manufacturing companies. Journal of Manufacturing Technology Management **31**(5), 955–976 (2020). https://doi.org/10.1108/JMTM-05-2018-0137
10. K., M., et al.: A survey (nlp) natural language processing and transactions on (nnl) neural networks and learning systems. E3S Web of Conferences **430**, 01148 (2023). https://doi.org/10.1051/e3sconf/202343001148
11. Mark, B.G., Rauch, E., Matt, D.T.: Worker assistance systems in manufacturing: A review of the state of the art and future directions. Journal of Manufacturing Systems **59**, 228–250 (2021). https://doi.org/10.1016/j.jmsy.2021.02.017
12. Schmidbauer, C., et al.: An empirical study on workers' preferences in human–robot task assignment in industrial assembly systems. IEEE Transactions on Human-Machine Systems **53**(2), 293–302 (2023). https://doi.org/10.1109/THMS.2022.3230667
13. Simões, B., et al.: Cross reality to enhance worker cognition in industrial assembly operations. The International Journal of Advanced Manufacturing Technology **105**(9), 3965–3978 (2019). https://doi.org/10.1007/s00170-019-03939-0
14. Winkler, R., Söllner, M., Leimeister, J.M.: Enhancing problem-solving skills with smart personal assistant technology. Computers & Education **165**, 104148 (2021). https://doi.org/10.1016/j.compedu.2021.104148
15. YouTube: Figure status update - openai speech-to-speech reasoning (4/19/2024), https://www.youtube.com/watch?v=Sq1QZB5baNw