CIRP Annals - Manufacturing Technology 73 (2024) 9–12

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: https://www.editorialmanager.com/CIRP/default.aspx

An LLM-based approach for enabling seamless Human-Robot collaboration in assembly

Christos Gkournelos, Christos Konstantinou, Sotiris Makris (2)*

Laboratory for Manufacturing Systems and Automation, Department of Mechanical Engineering and Aeronautics, University of Patras, Patras, Greece

A R T I C L E   I N F O

Article history:
Available online 22 May 2024

Keywords:
Large language models
Human robot collaboration
Manufacturing systems

A B S T R A C T

The complexity in the collaboration between humans and robots in smart manufacturing remains a significant challenge. This paper introduces an LLM-based manufacturing execution system enhancing Human-Robot Collaboration (HRC) in smart manufacturing. By leveraging Large Language Models (LLMs), the system provides a natural language interface for operators, integrates with Digital Twins for real-time data, and employs behavior-based control for robots. This integration facilitates intuitive interactions and rapid system programming, addressing communication complexities in HRC. The effectiveness of this approach is validated through two HRC assembly case studies, demonstrating significant improvements in collaboration and efficiency.

1. Introduction

In recent advancements in manufacturing [1], human robot collaboration (HRC) has been employed significantly as a means to increase shop floor flexibility. The need for producing a large variety of customizable products increases the required resilience of the production system, which in turn requires further advances in flexible collaborative manufacturing systems [2]. Human centered manufacturing emphasizes the collaboration of artificial intelligence (AI) agents with human workers while presenting social values. Despite the recent advances in recent years, there are still limitations in task planning, intuitive interaction, and programming [3].

Data architecture in HRC-based production systems often remains highly complex [4], necessitating employing advanced frameworks to handle and process large volumes of data. Moreover, making these systems user-friendly for non-specialists is still a significant challenge [5]. It's essential to develop systems with clear, intuitive interfaces and straightforward interactions to facilitate easy engagement. Striking the right balance between managing complex data and ensuring user accessibility is crucial in creating effective HRC systems that seamlessly combine technical sophistication with approachable interfaces.

Progress on AI [6], and especially on generative AI, has shown promising results, creating novel interactions approaches, with simpler scenarios [7]. Studies have shown that such AI models can efficiently adapt to complicated data structures transformed into their vectorized form [8], and by using their intrinsic knowledge, be able to respond to queries only by exploiting the nativeness of the natural language. This advantage of the generative AI, and especially Large Language Models (LLMs), can be interpreted as a middle stage assistant solution, having an agent-like behavior, by formulating plans and utilizing external tools [9], lowering the barrier of a collaboration between robots and humans.

The tendency of LLMs to perpetuate biases, coupled with their vulnerability to producing hallucinatory [10] or erroneous content [11], presents a significant challenge. It is crucial to strategically deploy self-supervised deep LMs in a manner that not only overcomes these technical hurdles but also enhances their utility in practical engineering applications.

This paper presents a novel approach for efficient and intuitive coordination and interaction of human and robotic resources in assembly tasks. Leveraging the extensive semantic knowledge of LLMs, the system is adept at identifying practical tasks for collaborative assembly between humans and robots. This approach enables easy programming for HRC assembly systems and enhances non-expert user interactions.

The system enables production engineers to effortlessly orchestrate the process schedule and allocate tasks to the resources using text commands. While a query command like "create a schedule for the final assembly of the e-motor with one robot and one worker", may elicit a coherent narrative response by an LLM, it lacks practical applicability in the context of assembly execution. This is because LLMs, in their standard form, are not inherently attuned to the manufacturing world. Robots follow precise commands, while human workers need clear, concise instructions.

This paper details a method for implementing an LLM application to extract relevant information for manufacturing, manage the task execution, and enable natural language interaction with human workers and engineers. This solution bridges the gap between the narrative capabilities of LLMs and the practical requirements of a physical, collaborative manufacturing environments and focuses on the creation of a "manufacturing reasoning" by an LLM.

* Corresponding author.
E-mail address: makris@lms.mech.upatras.gr (S. Makris).

https://doi.org/10.1016/j.cirp.2024.04.002

The paper is organized as follows. The Section 2 presents the systematic methodology and the implementation details of the system. In Section 3, the application of the proposed system in two industrial case studies is presented. Section 4 presents the evaluation of the system while in Section 5 the conclusions and future developments are reported.

2. LLM-based execution system for HRC

The proposed LLM-based manufacturing execution system is designed to integrate and enhance HRC in smart manufacturing environments. The system is characterized by several key features:

1. **System knowledge**: Access to manufacturing proprietary knowledge bases and data sources.
2. **Online information**: Integration with the Digital Twin (DT) of the system allowing the access of dynamically updated data. This feature is mandatory when there is a need to obtain and use the runtime state of the system.
3. **Scalability**: Modular architecture for allowing easy reconfiguration based on application requirements. The training of the AI components is performed independently to easily add new capabilities without compromising the performance of the existing ones.
4. **Explainability**: A transparent process that can be monitored by humans allowing a level of control in the observations and decisions.

The framework of the proposed system is illustrated in Fig. 1. The proposed LLM-based approach for HRC is based on an extendable set of modules, each serving a distinct function within the HRC environment. There are two primary categories of modules: a) Agents and b) Modules. The Agents are neural modules including LLMs, fine-tuned LMs, or any other AI enabled component with the objective to classify an action based on inputs. Modules are symbolic tools that implement action functions called by Agents. These modules are used for connecting external APIs, such as databases, robot controllers, etc.

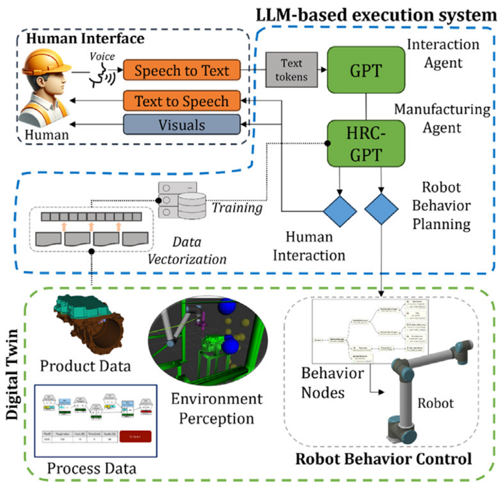

Fig. 1. The proposed LLM-based approach for HRC execution system.

The proposed system consists of one formating Agent with NLP capabilities, two language model (LM) Agents (*Interaction Agent* and *Manufacturing Agent*), and two action Modules (*Robot Behavior Planning, Human Interaction*).

2.1. Agents

Within the proposed LLM-based execution system for HRC, agents are integral neural modules tasked with system observation and interaction. The system incorporates two specialized agents: the "*Interaction Agent*" and the "*Manufacturing Agent*." Both agents are fine-tuned utilizing the chain-of-thought (CoT) [12] method, which is instrumental in minimizing hallucination issues. By defining a sequence of reasoning steps, this method narrows the response space, thereby focusing on targeted, relevant outputs.

The *Interaction Agent* is primarily engaged in handling the conversation with the users. It employs a 'plan-and-execute' strategy and is equipped with a conversational memory buffer, enhancing its ability to manage and sustain user interactions. To meet the natural interaction requirement, a series of tests were conducted to select the most suitable LLM for this agent. GPT-3.5 emerged as the preferred choice, attributed to its high accuracy in generating responses and optimal response time. This selected model has been fine-tuned using CoT, enabling it to effectively discern whether an input should be redirected to the Manufacturing Agent or addressed as a general query.

The *Manufacturing Agent* addresses questions related to the HRC assembly deployed in the system. A custom LLM, HRC-GPT, based on the gpt-4–0613 model [13], was specifically trained for this purpose. This agent is responsible for generating comprehensive production schedules or modifying existing tasks in the running assembly schedule. HRC-GPT possesses intrinsic knowledge of the production system, a capability enhanced through the integration of manufacturing data via vector augmentation and prompting techniques, as outlined in Section 2.2. Endowed with "manufacturing reasoning," HRC-GPT is adept at creating accurate assembly schedules and interfacing with the respective modules that manage resource control. This functionality underscores its pivotal role in the HRC process, enabling precise and efficient coordination of manufacturing activities.

2.2. Modules

The methodology proposed in this paper includes two main modules that serve as bridges to the physical world: the Human Interaction Module and the Robot Behavior Planning module.

The *Human Interaction* module is designed to facilitate natural communication between human operators and the system. It is equipped with Natural Language Processing (NLP) capabilities, including both speech-to-text and text-to-speech functionalities.

This module plays a dual role in the LLM-based system: firstly, as an input interface, enabling human operators to interact seamlessly with the system through natural language; and secondly, as an output interface, effectively communicating system-generated information back to the human users. Its design focuses on ensuring that interactions are intuitive and user-friendly, thereby enhancing the overall efficiency and effectiveness of the HRC process.

In managing robotic resources, the system adheres to the Behavior Trees (BTs) methodology [13]. This approach is particularly advantageous across various applications due to its structured manner of representing and controlling autonomous robots. The *Robot Behavior Planning* module plays a critical role in this process. It is tasked with receiving the BTs generated by the Manufacturing Agent and subsequently relaying them for execution. This methodology empowers the LLM system to produce precise outputs that are not only theoretically sound but also practically executable in the physical world. Through this integration, the LLM system transcends traditional boundaries, effectively bridging the gap between high-level planning and real-world implementation.

2.3. Manufacturing data and training

The training of the LLM for the specific HRC knowledge integration into our assembly system employs a comprehensive approach, utilizing vectorized representations of diverse data sources. As depicted in Fig. 2, the input manufacturing data is organized in several structured files. This data encompasses several critical components: process planning details, comprehensive resource information, geometric layout data, and a collection of primitive behaviors designed for robotic execution.

Resource data provides an extensive inventory of tools and equipment, outlining their constraints and operational limits. These data are considered crucial for facilitating improved resource allocation aiding the planning of task scheduling. Geometric layout data,

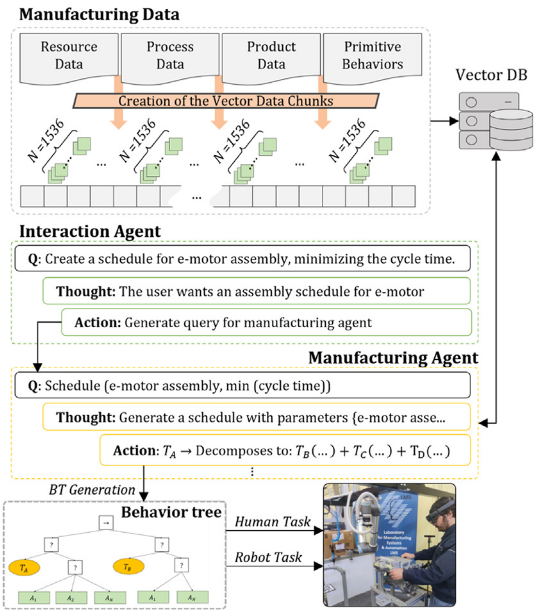

Fig. 2. The overall system architecture.

extracted from CAD and Digital Twin sources, precisely map the positions and orientations of all components within the manufacturing cell, ensuring spatial accuracy in planning. The process plan data delineates the sequence of assembly and disassembly tasks, while the primitive behaviors represent a suite of fundamental actions executable by robots.

To effectively incorporate this data into the LLM, it undergoes transformation into a vectorized format using an embedding model. Each data segment is processed to generate a semantic vector. This vectorization is performed using the text-embedding-ada-002 model [13], which produces a 1536-dimensional output matrix for each data segment. This matrix is used as a foundation for the creation of the HRC-GPT. The custom trained model employs a cosine similarity metric [14] to retrieve relevant data segments, which are conceptualized as neighboring vectors in the vector space. The system then generates a prompt that combines this retrieved data with the internal memory summary. This method in collation with the CoT prompt engineering enables the generation of contextually relevant and precise responses for the HRC system.

3. Case study

To validate and assess the effectiveness of the proposed LLM-based manufacturing execution system, a comprehensive testbed environment was established, featuring two distinct manufacturing scenarios. The first scenario, derived from the automotive industry, involves the intricate assembly of an electric motor and its inverter. The second scenario originates from the machinery industry, focusing on the assembly of an industrial air compressor unit. These cases were specifically chosen for their relevance in the HRC domain, particularly due to the flexibility they offer in task assignments. Such flexibility is crucial in evaluating the LLM's capabilities in schedule generation and task allocation.

Both assembly cases involve a UR10 collaborative robot and a human operator. Additionally, a variety of tools are provided to facilitate the flexible operation of the robot. These tools include a pneumatic suction gripper, two mechanical grippers, and four screwdrivers – two electric and two pneumatic. The tasks designed for this environment range from heavy picking and placing operations, suitable for the robot, to screwing tasks. Meanwhile, there are also picking and placing tasks that require collaboration of human operator with the robot. A significant proportion of the assembly tasks are designed to be versatile, capable of being performed by either the robot or the human operator. In the Table 1, the number of robot, human, and collaborative tasks are depicted. In this work two use cases of the system are investigated.

Table 1
Number of tasks in the industrial use cases.

Number of Tasks	T_{total}	T_{robot}	T_{human}	T_{collab}
E-motor	16	15	14	1
Air compressor	14	14	14	6

3.1. Task planning and system programming

The initial application of the LLM-based execution system in our case study involves transforming natural language inputs from production engineers into a detailed production schedule. This process is driven by specific Key Performance Indicators (KPIs). The primary objective of the system is to generate a schedule that not only adheres to the semantic rules embedded in the query but is also executable in the physical manufacturing environment. Fig. 3 shows an interaction example on a user query. The system employs a 'chain of thoughts' approach, enabling it to grasp the underlying meaning of the input query comprehensively. Consequently, the system's output is not just a theoretical plan but a practical, executable schedule that can be directly implemented in the physical manufacturing cell.

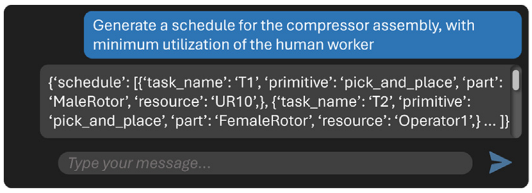

Fig. 3. HRC schedule generation from a simple language instruction.

3.2. Dynamic reconfiguration

The second use case for the LLM-based execution system involves its role in facilitating dynamic interaction with the human operator during execution phases. Central to this interaction is an innovative interface, named *CollabAI,* which is deployed on Augmented Reality (AR) glasses and serves as a critical communication link between the operator and the system (Fig. 4). This AR interface is designed to support both voice and text modalities. The LLM system provides real-time guidance and feedback. This level of interaction enhances the operator's situational awareness and decision-making capabilities and ensures that the manufacturing process remains flexible and responsive to real-time changes.

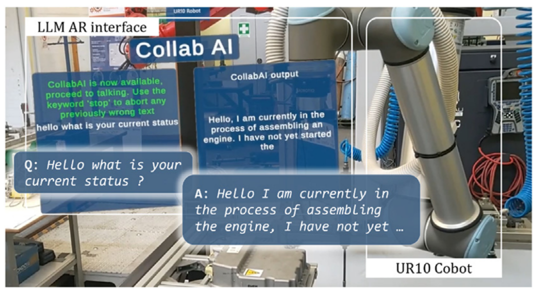

Fig. 4. Augmented reality interface of the LLM-system.

4. Experimental evaluation

To assess the performance of HRC-GPT, a dataset was created and tailored to the evaluation of natural language-based task planning in assembly cases. Derived from the environments and actions detailed in two industrial cases, the dataset is divided into three categories covering different task complexities and Key Performance Indicators (KPIs)

Simple dataset comprises data pertaining only to the suitability of resources for each task. Utilizing this dataset, the system's performance is gauged based on *Resource Utilization* as the sole KPI for process schedule selection.

Intermediate dataset is enriched with additional information on each task's process time by the resources. It allows for the analysis of two KPIs: *Resource Utilization* and *Assembly Cycle Time*.

Complex dataset: Incorporating all the information related to the assembly process, this is the complete dataset of each case. It enables the generation of schedules considering up to five criteria: Resource Utilization, Cycle Time, Non-adding Value Activities Time, Ergonomics, and Safety.

The experimental methodology was structured around these three dataset categories. Each use case was associated with a group of datasets, and an additional combined case was created by merging data from both the electric motor and the air compressor assembly cases.

A diverse set of input queries were crafted for the experiments. Initially, one query was designed for each dataset case and type. Subsequently, the general GPT-4 model was utilized to generate multiple variations of the initial query, validating performance across different phrasings. For the combined case, the queries from both the electric motor and air compressor assembly were employed. Starting from 30 distinct input queries (5 for simple, 10 for intermediate, and 15 for complex), our augmentation technique expanded this to a set of 1000 input queries, with 500 for each case (50 for simple, 100 for intermediate, and 350 for complex).

The output evaluation involved comparing the HRC-GPT's responses to the outputs of the AI-task planning tool [15], which served as a ground truth. For each input query, the task planning tool was configured accordingly. For the evaluation three metrics were employed:

- Format Accuracy (*FA*): This metric evaluates the HRC-GPT's responsiveness to deliver a structurally correct output.
- Executability (*EX*): This assesses the system's ability to deliver functionally correct schedule, ready for execution. "Correct" here refers to the accurate assignment of resources to tasks they can perform.
- Goal Accuracy (*GA*): This measures the correlation between the expected output schedule and the generated one, particularly in terms of KPI compliance.

The results of the experiments are presented in Table 2. Starting from the natural language, up to the generated schedule, the system demonstrated consistent structural correctness, evidenced by high FA across all datasets. EX indicated effective functional schedule generation, though with a slight decline in more complex scenarios. This indicates that LLM based system are not governed by causality but are the perfect candidate to correlate vast amounts of data, even unseen ones. The most notable challenge was observed in GA, particularly in more intricate tasks, reflecting difficulties in aligning with multiple KPIs. Notably, in approximately 70% of cases, the desired output was achieved with just one additional user input, highlighting the system's responsiveness and potential for user-guided optimization. Overall, HRC-GPT shows strong structural accuracy, but its functional correctness and multi-KPI optimization, especially in complex environments, present opportunities for further development and refinement.

Table 2
Evaluation of HRC-GPT.

	Electric Motor			Air Compressor			Combined		
Metrics	FA	EX	GA	FA	EX	GA	FA	EX	GA
Simple	1	1	0.9	1	0.9	0.9	1	1	1
Intermediate	1	0.8	0.9	1	0.9	0.6	1	0.8	0.8
Complex	1	1	0.8	1	0.8	0.6	1	0.8	1

5. Conclusions

This research establishes a robust framework using an LLM-based system as an orchestrator in manufacturing environments. The primary objective of this paper was to elucidate the concept of "manufacturing reasoning" and its necessity in leveraging LLM capabilities. The proposed solution showcased an end to end (from programming to execution) application of an LLM system in an actual HRC assembly station. The deployment and evaluation of this approach using real industrial data have highlighted the significant potential of this system in practical scenarios.

Future work will go deeper into refining training methodologies and prompt engineering strategies to improve the system's performance in complex reasoning involving multiple parameters. Additionally, another goal is a dataset creation using diverse manufacturing tasks, serving as a benchmark to assess and compare methodologies, further advancing in the field.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Christos Gkournelos: Conceptualization, Formal analysis, Software, Writing – original draft, Writing – review & editing. **Christos Konstantinou:** Software, Validation, Writing – original draft, Writing – review & editing. **Sotiris Makris:** Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

Acknowledgments

This research has been supported by the "CONVERGING Social industrial collaborative environments integrating AI, Big Data and Robotics for smart manufacturing (Grant Agreement: 101058521)".

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.cirp.2024.04.002.

References

[1] Chryssolouris G (2006) *Manufacturing Systems: Theory and Practice*, Springer.
[2] Makris S (2021) *Cooperating Robots for Flexible Manufacturing*, Springer International Publishing.
[3] Krüger J, Wang L, Verl A, Bauernhansl T, Carpanzano E, Makris S, Fleischer J, Reinhart G, Franke G, Pellegrinelli S (2017) Innovative Control of Assembly Systems and Lines. *CIRP Annals* 66:707–730.
[4] Wang L, Liu S, Liu H, Wang XV (2020) Overview of Human-Robot Collaboration in Manufacturing. In: *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing*, 15–58.
[5] Wang L, Gao R, Váncza J, Krüger J, Wang XV, Makris S, Chryssolouris G (2019) Symbiotic Human-Robot Collaborative Assembly. *CIRP Annals* 68:701–726.
[6] Chryssolouris G, Alexopoulos K, Arkouli Z (2023) *A Perspective on Artificial Intelligence in Manufacturing*, Springer.
[7] Zhang Z, Chai W, Wang J (2023) Mani-GPT: A Generative Model for Interactive Robotic Manipulation. *Procedia Computer Science* 226:149–156.
[8] Asai A, Min S, Zhong Z, Chen D (2023) Retrieval-Based Language Models and Applications. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, Tutorial Abstracts, , 41–46.
[9] Ruan J, Chen Y, Zhang B, Xu Z, Bao T, Du G, Shi S, Mao H, Zeng X, Zhao R (2023) TPTU: Task Planning and Tool Usage of Large Language Model-based AI Agents. *NeurIPS-2023 Workshop on Foundation Models for Decision Making*, arXiv preprint arXiv:2308.03427..
[10] Han Y, Liu C, Wang P (2023) A Comprehensive Survey on Vector Database: Storage and Retrieval Technique. *Challenge* . arXiv preprint arXiv:2310.11703.
[11] Bender EM, Koller A (2020) Climbing Towards NLU: On Meaning, Form, and Understanding In The Age of Data. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.
[12] Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E, Le Q, Zhou D (2022) Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems* 35:24824–24837.
[13] Cao Y, Lee G (2022) Behavior-Tree Embeddings for Robot Task-Level Knowledge. *2022* IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 12074–12080.
[14] Han J, Kamber M, Pei J (2012) Getting to Know Your Data. *Data Mining* : 39–82.
[15] Evangelou G, Dimitropoulos N, Michalos G, Makris S (2021) An approach for task and action planning in Human–Robot Collaborative cells using AI. *Procedia CIRP* 97:476–481.