



Leveraging the Power of Large Language Models to Drive Progress in the Manufacturing Industry

Feng Shi¹, Yongjin Zhang², Chongxiao Qu², Changjun Fan^{2,*}, Jinqi Chu², Lei Jin², Shuo Liu²

¹Cethik Group Co., Ltd., Hangzhou, Zhejiang, 311000, China

²The 52nd Research Institute of China Electronics Technology Group Corporation, Hangzhou, 311000, China

*Corresponding Email: cethiker@126.com

Abstract. In this paper, we delve into the transformative role of large language models (LLMs) in the evolving digital economy, particularly within the manufacturing industry. We begin by demystifying the fundamental principles of LLMs, aiming to deepen the reader's comprehension. Subsequently, we explore the main strategies for adapting LLMs and integrating them into manufacturing processes, scrutinizing both their benefits and limitations. Additionally, we assess the multifaceted impacts of LLMs on various phases of manufacturing. The paper culminates in a forward-looking analysis, highlighting four emergent trends that signify the growing influence of LLMs in revolutionizing the manufacturing industry.

Keywords: large language models, manufacturing industry, Transformer, ChatGPT.

1 Introduction

Manufacturing is a vital and indispensable element in the high-quality development of the economy. The advent of 5G+, cloud computing, artificial intelligence (AI), digital twins, and industrial internet technologies has propelled the digital transformation of manufacturing into a new era [1]. The widespread integration of the industrial internet and AI in production and management processes is now a pivotal force in elevating manufacturing enterprises, playing a critical role in fostering high-quality economic progress. A defining feature of this new phase in the digital economy, driven by the industrial internet, is the seamless fusion of digital and physical systems. This convergence is becoming a key factor in reshaping the global competitive landscape of manufacturing. As we edge closer to the Industry 4.0 epoch, the urgency for digital transformation in manufacturing intensifies [2]. However, industrial enterprises face the absence of a universal technological solution to tackle the varied and complex challenges present in different manufacturing environments. In this context, the rise of large language models (LLMs) offers a beacon of hope, potentially providing innovative solutions to these multifaceted challenges.

© The Author(s) 2024

K. Elbagory et al. (eds.), *Proceedings of the 9th International Conference on Financial Innovation and Economic Development (ICFIED 2024)*, Advances in Economics, Business and Management Research 281,

https://doi.org/10.2991/978-94-6463-408-2_15

LLMs have created a significant wave in the natural language processing field. The 21st century witnessed its groundbreaking progress, including the advent of recurrent neural networks (RNN), the innovation in word embedding techniques, and the implementation of the word2vec algorithm. The introduction of the Transformer architecture in 2017 set the stage for significant strides, with OpenAI's GPT series and Google's BERT as its two most notable implementations. The debut of ChatGPT in late 2022 and the multimodal prowess of GPT-4 have elevated its global impact [3]. Besides, leading Chinese corporations and academic institutions are making headway in this arena, employing Transformer-based models in diverse areas like machine translation, voice recognition, customer service, algorithmic recommendations, and even programming. There's a growing belief that these models might evolve into ubiquitous platforms on par with computer operating systems, potentially revolutionizing people's daily work and lives. Essentially, LLMs leverage deep learning to process and interpret large volumes of text data, fostering content comprehension and creation. They have become indispensable in various industries, including manufacturing, to serve as advanced AI tools that interpret complex processes and provide valuable insights. After being tailored to meet specific needs, LLMs' application in manufacturing is expanding, with wide uses in various aspects like predictive maintenance, defect detection, and production planning, fueled by technological progress and reduced implementation costs.

This paper is dedicated to exploring the potential of leveraging LLMs to drive progress in the manufacturing industry. Initially, it introduces the basic principles of LLMs, aiming to enhance readers' understanding of them. Following this, it examines the fundamental approaches of tailoring LLMs and integrating them into the manufacturing process and analyzes their pros and cons. It then discusses the impact of LLMs on various phases of manufacturing. And, it culminates with an exploration of four basic future trends in empowering manufacturing with LLMs.

2 Technical principles of LLMs

The foundation of large language models is the Transformer architecture, with the "T" in GPT and BERT referring to this [4]. China's top AI models like 'ERNIE Bot' and 'iFLYTEK Spark', as well as international giants' models like PaLM-E and LLaMA, are also based on Transformer. While other models like generative adversarial networks (GAN) also aid in training, Transformer is the mainstay in supporting mainstream large models. It differs from other neural network models like GAN and RNN, using a multi-head attention mechanism to process input and output texts, as shown in Fig. 1.

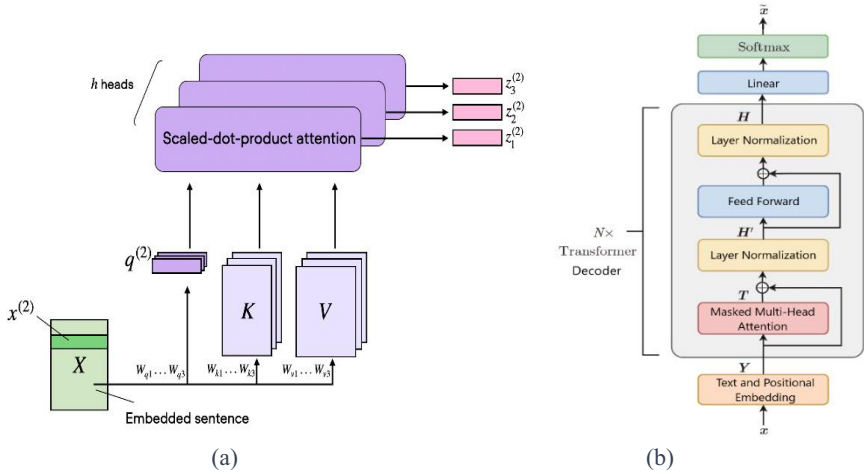


Fig. 1. (a) The multi-head attention mechanism; (b) The Transformer decoder adopted by ChatGPT

The function of Transformer fundamentally involves calculating the predicted probability for each position's output to complete the entire output sequence. This process primarily involves a decoder that processes the input text, converting the corresponding output into probabilities for the subsequent output token. Each token produced by the model signifies the most probable word or character for that specific position. As each word or character is generated, it is appended to the output sequence, and the model proceeds to predict the next token's probability, repeating this cycle until the final output is formed. In the absence of supplementary technologies, the Transformer's objective is to maximize probability as per its evaluation criteria, but this does not inherently ensure complete accuracy or truthfulness. This predictive mechanism involves compressing vast learned data into smaller datasets, inevitably leading to some loss of information. The inherent probabilistic filtering and selection processes in the model can increase the chances of generating incorrect responses. The accuracy and relevance of the responses are influenced by both the quality of the training data and the nature of the queries posed to the model.

The Transformer architecture, known for its well-structured, straightforward, and efficient computational design, is adept at handling complex tasks through the integration of large-scale modules. It contains four key vector groups within its "attention heads": Query (Q), Key (K), Value (V), and Weight (W). The architecture's attention layer dynamically synthesizes information by facilitating interactions between Query and Key-Value pairs. It projects the input sequence into these vectors and computes the similarity between Q and K. This similarity score is then modulated by W to influence V. This mechanism of Transformer can be analogized to a real-world scenario: imagine a girl using this mechanism to choose the most appropriate skirt from a collection. Here, Q represents her desired skirt attributes (fabric, color, style, brand, etc.), while K encompasses these features for each skirt. The more the features in Q and K keep in align-

ment, the higher the similarity score (indicating the skirt closely matches her preference). The relevance of each feature is weighted according to the continuously optimized coefficients in W , ultimately guiding her to select the skirt she prefers most. This complex mechanism contributes to the substantial size of models like GPT-2 and GPT-3. For instance, GPT-2 is composed of 48 decoder layers with 1.5 billion parameters, whereas GPT-3 includes 96 decoder layers and boasts 17.5 billion parameters [5]. This expansive scale enables LLMs to process intricate inputs effectively and maintain robust, adaptable learning capabilities. However, it also presents challenges for conventional algorithm interpreters and external algorithm transparency technologies to function effectively, given the complexity and scale of these models.

The Transformer architecture exhibits the potential for self-explanation. While the attention mechanism alone is not equivalent to an explanation, it can be translated into credible and meaningful interpretations. The core vectors, Q , K , and V , derived from word vectors and corresponding matrices, play a crucial role in determining the model's output. Despite the vectors being initially random and meaningless, a meaningful attention structure can be designed by aligning Q with specific query features and K with corresponding features of the target. Understanding how similarity and attention weights are calculated and applied within the model can reveal where the model's focus lies, which offers the potential for a human-comprehensible, visual explanation of the model's workings, even when Q and K are not directly interpretable.

To improve the precision and applicability of LLMs, advanced techniques like nucleus sampling, temperature sampling, and chain-of-thought reasoning are integrated [6]. These methods significantly enhance the models' ability to grasp deeper relationships and causality among topics, facilitating more insightful reasoning and decision-making. The flexibility of these techniques in adjusting outputs lends a certain level of explainability to LLMs. This robust foundation enables developers to initially pre-train these models in general reading comprehension and adaptability. Such a preliminary training regime equips the models with a versatile skill set, paving the way for subsequent, more focused training and optimization in specific fields, such as manufacturing.

3 Custom solutions to manufacturing based on tailored LLMs

At present, the training of general-purpose LLMs relies heavily on vast quantities of unstructured text sourced from the internet, endowing these models with abilities like reading comprehension, text summarization, and text generation. An exciting new chapter is emerging for LLMs as they venture into the complex realm of manufacturing. However, adapting existing LLMs to the specific field of manufacturing demands targeted customization. Integrating advanced tailored LLMs into the manufacturing industry opens up a world of possibilities, marking the advent of a new era in intelligent manufacturing.

Tailoring LLMs for the manufacturing industry is a multi-faceted journey that begins with a comprehensive understanding of the industry's unique challenges. Key areas such as production bottlenecks, gaps in quality control, and inefficiencies in supply

chains are identified as primary targets for LLM intervention. This initial step is crucial for setting the direction and focus of the LLMs-based application to the manufacturing industry. The second step involves extensive collection and meticulous refinement of manufacturing-specific data. This data, ranging from sensor readings and production logs to maintenance reports, forms the core knowledge base for the tailored LLMs, equipping them with a detailed understanding of the complex interplay between machinery and materials. This rich dataset is instrumental in enabling LLMs to make informed, relevant decisions and predictions in the manufacturing context. Thirdly, selecting the appropriate technological tools is vital for the success of LLMs in manufacturing. Specialized AI frameworks designed to handle time-series data, along with advanced natural language processing tools capable of interpreting production reports, are key components in ensuring the LLMs' efficacy. The right combination of tools empowers LLMs to process and understand the intricate details of manufacturing data effectively. The fourth step is the creation and continual refinement of tailored LLM models, designed specifically to address particular tasks within the manufacturing process. This involves an iterative cycle of training, testing, and refining the models to enhance their accuracy and utility. This process of continuous improvement is crucial for LLMs to remain relevant and effective in the ever-evolving manufacturing landscape. Furthermore, these intelligent LLMs should be designed to seamlessly integrate into existing manufacturing systems, harnessing real-time data from sensors and machinery and fully unleashing their potential. This ongoing interaction creates a dynamic feedback loop, enabling LLMs to continually learn from and adapt to complex and real-time changing manufacturing scenarios. Such real-time data integration transforms LLMs into an active, responsive participant in the manufacturing process, capable of making immediate adjustments and predictions based on the latest information. Overall, integrating LLMs into manufacturing is one of continuous evolution and adaptation. We should consider not only the development of sophisticated models but also their seamless incorporation into the complex and dynamic environment of manufacturing.

In manufacturing, custom LLMs offer substantial benefits but also pose challenges. They can improve operational efficiency through task automation, production optimization, and failure prediction while ensuring product quality and reducing waste via anomaly detection and predictive maintenance. Tailored LLMs can streamline supply chain management, aid in resource allocation, and support worker decision-making with data-driven insights. They may also promote sustainable practices and enhance customer satisfaction through personalized production strategies, leading to cost savings and improved production times. However, tailoring and integrating LLMs involves considerable time, financial investment, and expertise. Challenges include acquiring production data for training, adapting existing systems for AI integration, and managing operational disruptions. Continuous maintenance and updates are also necessary for effectiveness. Ethical considerations, such as job deprivation and data security, require careful weighing, prudent consideration, and a focus on augmenting human skills. Besides, responsible AI deployment and documentation for knowledge advancement in the field should also be needed.

4 Impact of LLMs on various phases of manufacturing

LLMs are spearheading the third wave of digital-physical integration, characterized by enhanced intelligence. This integration will significantly influence the manufacturing industry, spanning research and design, production processes, quality management, operational control, marketing services, organizational collaboration, and business management.

In the phase of research and design, LLMs are set to transform conventional R&D paradigms. If leading electric vehicle companies, such as Tesla and BYD, are planning to develop a new type of electric car. In this endeavor, LLMs can play a significant role by significantly enhancing R&D efficiency. LLMs can be utilized to analyze and organize vast amounts of consumer feedback, market research reports, and historical data to determine consumer expectations and preferences for new electric vehicles. LLMs are capable of summarizing and extracting lessons from the successes and failures of past projects. Based on the compiled data, LLMs can offer innovative design suggestions. For instance, they might recommend the use of specific materials in the car's interior to improve energy efficiency or enhance passenger comfort. LLMs also can generate detailed technical documentation and design blueprints, aiding engineers in understanding and implementing these suggestions. When faced with technical challenges during this process, LLMs can provide solutions and optimization recommendations, such as improvements in battery technology or aerodynamic design. Additionally, LLMs can facilitate communication between the design team and other departments (like marketing and supply chain), ensuring that the design aligns with market demands and production realities. They can also automate the creation of project reports and briefing materials, thus increasing team collaboration efficiency.

In the phase of manufacturing production, LLMs are instrumental in advancing product innovation in sectors like smart cars, robotics, chips, and apparel. For instance, leading smartphone manufacturers such as Xiaomi and Samsung can leverage LLMs to significantly boost production efficiency and maintain product quality while developing new product models. LLMs play a key role in analyzing and organizing essential production data, including machine operation durations, raw material usage, and product defect rates. They excel in pinpointing process bottlenecks and inefficiencies during production, based on which LLMs offer specific recommendations for process improvements. This might involve adjusting machine settings or modifying raw material supply schedules to reduce waste. Additionally, LLMs aid in orchestrating production schedules, ensuring optimal operation sequences on the assembly line. They scrutinize quality control data to preemptively detect and address potential defect causes. LLMs are also adept at forecasting equipment malfunctions, suggesting timely maintenance to avert unplanned stoppages. Lastly, these models streamline documentation by automating the creation of production reports, quality assurance records, and maintenance logs, thus significantly improving document management efficiency in the manufacturing process.

In the phase of equipment operation and maintenance, LLMs significantly elevate the proficiency of traditional models, endowing them with advanced understanding capabilities. Take the power industry as an example: when drones gather data on electrical

equipment in mountainous regions, conventional models might simply flag "non-standard cotter pins". However, advanced LLMs, such as GPT-4, with their multimodal image recognition abilities, can offer more precise assessments. For instance, they might identify specific problems with bolts and cotter pins on a red-painted insulator located near a highway, and recommend immediate on-site inspection and repair. This enhanced level of detail and actionable insight provided by LLMs greatly improves the efficiency and accuracy of equipment maintenance processes.

5 Future trends of LLM application in manufacturing

As a new tool of productivity, LLMs are destined to extend deeply from the content domain (like text-to-text and text-to-image generation) into the tangible production sector, sparking a new efficiency revolution in various aspects of manufacturing and accelerating the industry's shift towards intelligence.

Firstly, driving software upgrades will be a primary way through which LLMs empower the manufacturing industry. Industrial software, crucial to digital transformation, can be revolutionized by LLMs in development approaches, interaction methods, usage procedures, and business models. LLMs can significantly upgrade various software systems across R&D, management, production, and post-service, unlocking new possibilities for future advancements. The latest AI coding platforms, powered by LLMs, exhibit exceptional code understanding and generation skills. They support vital functions like code completion, test unit creation, code explanation, and debugging. The emergence of Model as a Service (MaaS) is shifting the focus to a model-centric development approach, simplifying industrial software development and boosting efficiency. An early 2023 assessment by CSDN noted that GPT-4's software programming capabilities are comparable to a domestic software engineer earning 30,000 ¥ monthly, equivalent to a Google L3 engineer with an annual salary of 180,000\$. In a U.S. software job experiment, an engineer with four years of experience, using AI tools, achieved a software development efficiency five times higher than an engineer with nineteen years of experience. This illustrates the significant impact LLMs have on enhancing software development productivity in the manufacturing sector.

Secondly, bridging data flow gaps will be an important value that LLMs bring to the manufacturing industry. Every advancement in human-computer interaction heralds a significant industry shift, and LLMs are at the forefront of this transformation. They are ushering in a revolution where natural language becomes the primary mode of interaction, drastically changing how people engage with the environment. This shift is set to have a profound impact on the future industrial landscape. The essence of digitalization in manufacturing is to harness the automatic flow of data to manage the uncertainty of complex systems. It's about delivering the right data, in the right format, at the right time, to the right recipients, whether people or machines, thereby improving resource allocation efficiency. However, many enterprises face the challenge of data flow gaps across various segments, often requiring engineers to create multiple process and flow software solutions. LLMs offer a novel solution to this issue. Leveraging their

natural language interaction capabilities, they present new approaches for software development and interaction. This enables real-time, omnipresent connections within manufacturing enterprises and across the industry. By lowering the development barriers for process and flow software, LLMs enhance efficiency and bridge numerous gaps in enterprise data flow, paving the way for more streamlined and effective manufacturing processes.

Thirdly, penetrating control segments will be a pivotal measure of how LLMs empower manufacturing. The primary value of LLMs in the manufacturing industry lies in their ability to infiltrate production control segments, beyond just marketing and management roles. The versatility and adaptability of LLMs, coupled with the innovative 'pre-training + fine-tuning' development approach, equip them to enhance various manufacturing phases, including R&D, production processes, operations, quality control, sales, customer service, and organizational collaboration. A critical aspect of this integration is the entry of LLMs into core production control systems like PLC, MES, SCADA, etc. This move significantly boosts the intelligence of production processes. For example, in SCADA systems, the strategy involves using LLMs' programming interfaces and industry-specific libraries to create industrial logic code, such as interaction protocols, modeling, and SQL development. This code is then automatically integrated into industrial software, with the model optimized based on feedback from closed-loop results. This step marks a significant milestone in applying LLMs in manufacturing, indicating their deep integration and impact on the industry's core operational systems.

Moreover, the synergy between large and small models represents a significant trend in the role of LLMs in enhancing manufacturing. While large models require specific application scenarios, they are not yet fully equipped to address practical issues across diverse fields. A critical development in industrial application will be the collaboration and coordination between general and specialized, open-source and proprietary models, alongside existing software and hardware systems. This stage is vital for their industrial implementation. In this context, AI Agents, functioning as a primary medium for high-level collaboration between large and small models, are emerging as new tools of production. An AI Agent, typically built on LLMs, is designed to autonomously perform specific tasks using various tools. These agents, working in concert with other models and external software, are adept at managing complex real-world tasks. Future AI Agents are envisioned to comprise perception systems, control systems, and execution systems. They will not only have generative abilities but will also excel in task comprehension, breakdown, scheduling, execution planning, and chain coordination. This evolution marks a significant stride in the practical application of AI in the manufacturing industry, enhancing efficiency and innovation.

6 Conclusion

The development of large language model technology offers tremendous opportunities for transformative change across various industries, with the potential to address challenges in traditional sectors like manufacturing. This article is dedicated to exploring

the potential of leveraging the power of LLMs to drive progress in the manufacturing industry. It began by unveiling the fundamental principles of LLMs, deepening the reader's understanding of these models. The article then examined the main strategies for tailoring LLMs and integrating them into the manufacturing process, carefully evaluating their benefits and limitations. Building on this, it also assessed the specific impacts of LLMs on different phases of manufacturing. The article concluded with a forward-looking analysis, presenting four emerging trends in the application of LLMs to the manufacturing industry.

Reference

1. Wang, Lihui. "From intelligence science to intelligent manufacturing." *Engineering* 5.4 (2019): 615-618.
2. Tao, Fei, et al. "Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: Correlation and comparison." *Engineering* 5.4 (2019): 653-661.
3. Liu, Yiheng, et al. "Summary of chatgpt-related research and perspective towards the future of large language models." *Meta-Radiology* (2023): 100017.
4. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
5. Dale, Robert. "GPT-3: What's it good for?" *Natural Language Engineering* 27.1 (2021): 113-118.
6. Lyu, Qing, et al. "Faithful chain-of-thought reasoning." *arXiv preprint arXiv:2301.13379* (2023).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

