Large Language Models (LLMs) for Smart Manufacturing and Industry X.0

Baptista, Marcia L.; Yue, Nan; Manjurul Islam, M. M.; Prendinger, Helmut

DOI
10.1007/978-3-031-80154-9_5

Publication date
2025

Document Version
Final published version

Published in
Springer Series in Advanced Manufacturing

Citation (APA)
Baptista, M. L., Yue, N., Manjurul Islam, M. M., & Prendinger, H. (2025). Large Language Models (LLMs) for Smart Manufacturing and Industry X.0. In *Springer Series in Advanced Manufacturing* (pp. 97-119). (Springer Series in Advanced Manufacturing; Vol. Part F138). Springer Nature. https://doi.org/10.1007/978-3-031-80154-9_5

Springer Series in Advanced Manufacturing

M. M. Manjurul Islam
Marcia L. Baptista
Faisal Tariq *Editors*

Artificial Intelligence for Smart Manufacturing and Industry X.0

Springer

Springer Series in Advanced Manufacturing

Series Editor

Duc Truong Pham, University of Birmingham, Birmingham, UK

The **Springer Series in Advanced Manufacturing** includes advanced textbooks, research monographs, edited works and conference proceedings covering all major subjects in the field of advanced manufacturing.
The following is a non-exclusive list of subjects relevant to the series:

1. Manufacturing processes and operations (material processing; assembly; test and inspection; packaging and shipping).
2. Manufacturing product and process design (product design; product data management; product development; manufacturing system planning).
3. Enterprise management (product life cycle management; production planning and control; quality management).

Emphasis will be placed on novel material of topical interest (for example, books on nanomanufacturing) as well as new treatments of more traditional areas.

As advanced manufacturing usually involves extensive use of information and communication technology (ICT), books dealing with advanced ICT tools for advanced manufacturing are also of interest to the Series.

Springer and Professor Pham welcome book ideas from authors. Potential authors who wish to submit a book proposal should contact Anthony Doyle, Executive Editor, Springer, e-mail: anthony.doyle@springer.com.

INDEXED IN SCOPUS

M. M. Manjurul Islam · Marcia L. Baptista ·
Faisal Tariq
Editors

Artificial Intelligence for Smart Manufacturing and Industry X.0

Springer

Editors
M. M. Manjurul Islam
School of Computing, Engineering
and Intelligent Systems
Intelligent Systems Research Centre
Ulster University
Londonderry, UK

Marcia L. Baptista
NOVA Information Management School
(NOVA IMS)
Universidade Nova de Lisboa
Lisbon, Portugal

Faisal Tariq
James Watt School of Engineering
University of Glasgow
Glasgow, UK

ISSN 1860-5168 ISSN 2196-1735 (electronic)
Springer Series in Advanced Manufacturing
ISBN 978-3-031-80153-2 ISBN 978-3-031-80154-9 (eBook)
https://doi.org/10.1007/978-3-031-80154-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Large Language Models (LLMs) for Smart Manufacturing and Industry X.0

Marcia L. Baptista, Nan Yue, M. M. Manjurul Islam, and Helmut Prendinger

Abstract The manufacturing industry is rapidly changing, creating a growing demand for more intelligent and adaptive systems. With recent developments in artificial intelligence, especially with the onset of large language models (LLMs) such as ChatGPT, new opportunities have emerged for companies to increase their productivity and maximize revenue. In a competitive environment, businesses must constantly innovate to stay ahead. To support innovative and competitive organizations, LLMs can analyze large amounts of data to identify trends and optimize processes. In addition, the industry faces a labor shortage, particularly in roles that require specialized skills. LLMs can fill this gap by providing real-time assistance and training. This knowledge transfer could help less experienced workers perform their tasks more effectively. Regulatory compliance is increasingly imperative in manufacturing, and LLMs can help ensure adherence to safety standards and regulatory requirements. LLMs can address these and other challenges by using their capabilities in data processing, natural language understanding, and predictive analytics. In this chapter, we explain the fundamental concepts behind LLM techniques and how to use them in a smart manufacturing environment such as Industry X.0. We discuss the challenges and future trends of LLMs in different industrial fields. We also highlight the need for LLM frameworks that can guarantee data privacy, security, and ethical usage.

M. L. Baptista (✉)
NOVA Information Management School, Universidade NOVA de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal
e-mail: m.baptista@novaims.unl.pt

N. Yue
TU Delft, Kluyverweg 1, 2629 HS Delft, The Netherlands
e-mail: n.yue@tudelft.nl

M. M. Manjurul Islam
Intelligent Systems Research Centre, Ulster University, Londonderry BT48 7JL, UK
e-mail: m.islam@ulster.ac.uk; manjurul.islam@ieee.org

H. Prendinger
National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
e-mail: helmut@nii.ac.jp

M. M. M. Islam et al. (eds.), *Artificial Intelligence for Smart Manufacturing and Industry X.0*, Springer Series in Advanced Manufacturing,
https://doi.org/10.1007/978-3-031-80154-9_5

Keywords Natural language processing · Large Language Models (LLMs) · Chat-GPT · Smart manufacturing · Industry X.0

1 Introduction

In the future, many industrial and manufacturing changes are expected to occur, mostly driven by new technologies. Advanced **robotics** will become an integral part of various industries, performing millimetric precision tasks and sometimes showing "superhuman efficiency" [8]. The demand for **specialized labor** will increase [15], requiring workers to have expertise in fields such as programming and **data analysis**. Compliance **requirements** will become **stricter**, necessitating adherence to safety, quality, and ethical standards [6]. Continuous **training** will be essential, with dynamic and interactive methods such as **virtual reality** (VR), **augmented reality** (AR), and individualized **learning platforms** becoming the norm [68]. The goal is to make this **transition** to Industry X.0 as **smooth** and **seamless** as possible for everyone involved.

To achieve more productive, efficient and innovative industries, we will necessarily need **large language models** (LLMs). **LLMs** such as **Chat-GPT** have revolutionized **Natural Language Processing** (NLP) by allowing a better understanding and generation of human language [22]. An LLM is an artificial intelligence (AI) model trained with an extremely large **amount of text** that can generate natural language output in response to prompt interaction [32]. The developments of LLMs have had considerable impacts on many **language tasks** in manufacturing, from real-time troubleshooting and support to collaborative design [47, 51]. Here, the idea is to let human creativity and the unique capabilities of machines work effortlessly **together**.

There are **many areas** of manufacturing that could benefit from having LLM systems. We illustrate these perspectives in Fig. 1. The diagram describes the various ways in which LLMs can be used to improve manufacturing processes. First, LLMs can help to generate new **design** ideas and optimize design through simulation. In addition, the collection of real-time information can help LLMs improve operational **processes**. The predictive and coding abilities of LLMs can play an important role in contributing to **maintenance** improvement and **quality control** as well as **supply chain** management. The conversational capabilities of LLM also facilitate employee **training** and knowledge transfer. Their ability to process large amounts of data can be used to ensure that industrial processes **comply** with regulations and standards.

In general, LLMs can **integrate** positively across these manufacturing functions. This chapter describes these functions **in more detail** and how they can be accomplished in practice. We will explore the role of LLMs in the manufacturing sector and advances to Industry X.0.

The remainder of the chapter is organized as follows. In Sect. 2 we start by describing in more detail the current and expected future impact of LLM in manufacturing. Section 3 with a revision of the field of natural language processing (NLP).

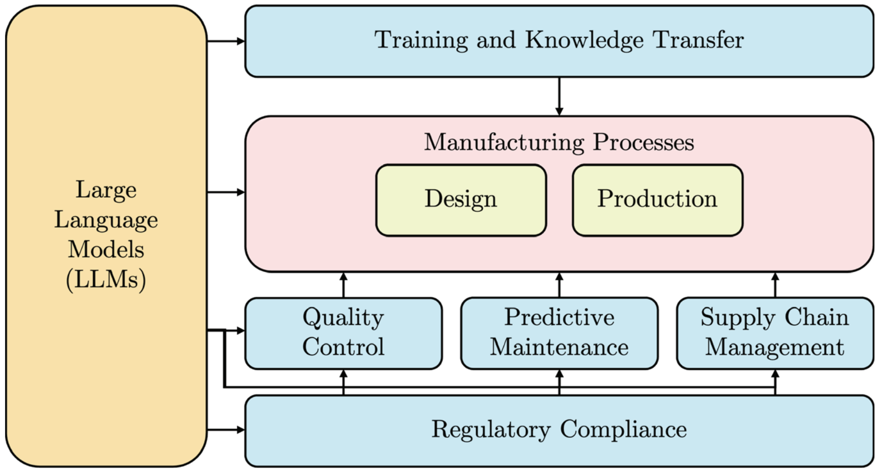


Fig. 1 Large Language Models (LLMs) in the manufacturing environment

Section 4 introduces large language models (LLMs) as well as its fundamental concepts. Section 5 discusses the applications of LLMs in Industry X.0. We discuss the challenges and ethical considerations of these methodologies and tools in Sect. 6. The latest and future trends are discussed in Sect. 7. Section 8 concludes the chapter with a reflection on the topic.

2 LLMs in Manufacturing

The **history** of large language models (LLMs) in manufacturing began in the 1950s with early NLP systems aimed at **language translation** [69]. These early rule-based systems were limited, but set the foundation for future AI developments. In the late **1980s** and 1990s, the introduction of statistical methods improved **data analysis** and pattern recognition. During this period, **manufacturing industries** started using machine learning for **predictive maintenance** and **quality control** [53], although these models were no LLMs.

The **2010s** saw the revolution of deep learning and neural networks, eventually leading to the development of highly sophisticated models like OpenAI's GPT series. Initially used for **language tasks**, LLMs quickly showed potential in other areas. In the late 2010s and early 2020s, LLMs started to be integrated into manufacturing processes [99]. **LLMs** are used to enhance various **dimensions** of **manufacturing**, from process automation and maintenance to machine control, training, supply chain, compliance and design. We categorize the impact of LLMs in **seven dimensions**, as illustrated in Fig. 2:

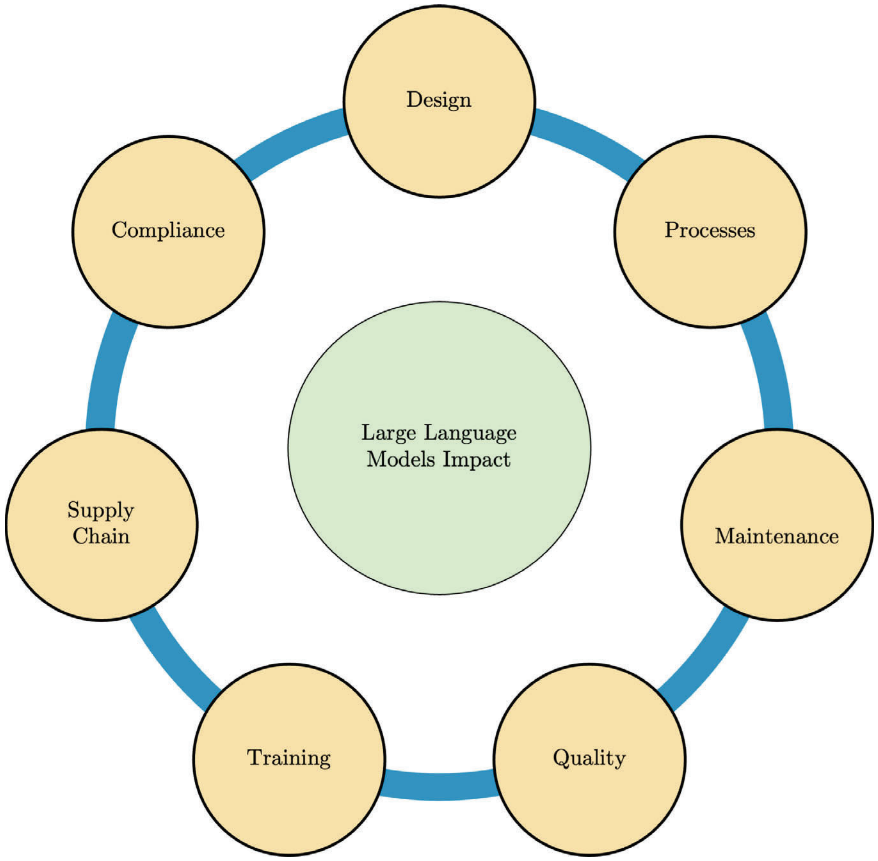


Fig. 2 Application of Large Language Models to Industry X.0

1. **Design**: LLMs assist in the design phase by offering suggestions and automating parts of the design process. LLMs can also be used in prototyping.
2. **Processes**: LLMs can analyze large amounts of data to identify inefficiencies and suggest improvements of operational processes.
3. **Maintenance**: by analyzing historical and real-time data, LLMs can predict equipment failures before they occur and enable predictive maintenance.
4. **Quality**: LLMs can assist in quality assurance by monitoring production data and detecting anomalies or defects, ensuring high standards.
5. **Training**: we can provide real-time troubleshooting and training to workers using LLM technology. LLMs can help in understanding complex tasks and procedures. This is particularly useful for new employees and technologies.
6. **Supply Chain**: we can use LLMs to optimize supply chain operations by predicting demand, managing inventory, and ensuring timely procurement of materials.

7. **Compliance**: LLMs help ensure compliance with industry regulations by monitoring processes, generating necessary documentation, and providing alerts when deviations occur.

The potential for innovation is **game changing**, making the integration of LLMs a **compelling** topic for anyone interested in the **future** of manufacturing technology. We explain each of these seven dimensions **hereafter**, providing **examples** of current and future applications, whenever necessary for better understanding.

Manufacturing design is a well-established field that is used for bringing products from idea to creation [55]. This field of study can be improved with LLMs across **different aspects** [62]. The LLMs can be used as repositories of information to generate new products and services. This technology can be used to **automate** the generative creation of new concepts, evaluate the performance of different concepts, and produce documentation. For example, Brahmavar et al. [10] investigate how to use LLMs and prompt engineering to generate pharmacological compounds. The authors show through several **examples** that this kind of discovery task can be automated with LLMs.

In the **future**, we will use LLMs in our design tasks in ways that are currently **unexplored**. For example, consider the challenge of reducing waste in the production of electronic components. An LLM could analyze thousands of **patents** related to electronic manufacturing. It could also have access to **scientific articles** on material science. We could provide the LLM system with **industry reports** on production efficiency. By synthesizing this information, the LLM could propose a novel recycling process for defective components. It could suggest using a specific polymer identified in the material science literature that can be easily separated from metals based on industry reports. Additionally, it might discover a patent technique, such as a method for 3D printing circuit boards with minimal waste. This example could soon become a **daily reality** for manufacturers. By automating routine tasks, engineers and designers can better concentrate on creativity and strategy, improving productivity, and fostering innovation in their specific industry. However, there are challenges, such as having the best data and using the LLM system most effectively.

The continuous improvement of business and operational **processes** is important to the overall efficiency of the **smart factory** [54]. In industrial settings, LLM systems can improve both manufacturing and business processes. In terms of manufacturing processes, LLMs can help automate routine tasks and automate robotic processes. Decision support systems based on LLMs can perform real-time analytics for better strategic and business decisions. Energy is also a central concern for the smart factory. Here, LLMs can improve energy efficiency by finding optimized consumption patterns. These advancements lead to increased efficiency, cost savings, and innovation. As an example, consider the work of Mayet [65] who introduced GAIA, an AI assistant for intelligent accelerator operations. The system helped human operators deal with a complex particle accelerator. Cybersecurity issues might need to be reviewed in these industrial settings [39]. Several other studies have investigated the importance of having conversational agents in an industrial setting [16, 23, 31, 35, 46]. Their findings were **positive**, but some of them highlighted concerns about bias, misinformation, overreliance, privacy, and also cybersecurity.

In the **near future**, the smart factory will operate with **automated processes** powered by LLMs. The scenario is one of personalization and **custom** orders. So, when a new order is received, the LLM interprets the customer requirements and translates them into precise production instructions. Autonomous robots and machines, guided by LLM, handle raw materials, assemble components, and manage the production line with **minimal** human intervention. The LLM also generates real-time reports, providing actionable insights for decision-makers. Robust cybersecurity measures, informed by the LLMs, safeguard the factory's data and operations.

Predictive maintenance is a proactive maintenance approach based on data analytics and machine learning. By analyzing **data** from **sensors**, maintenance **logs**, and **historical** performance, predictive maintenance algorithms can identify warning signs and recommend corrective actions. Zonta et al. [113] give an overview of predictive maintenance in Industry X.0. Another review is by Achouch et al. [2]. LLMs offer possibilities for analyzing unstructured data, such as maintenance logs, and extracting insights. Machines can help identify degradation patterns and ultimately predict equipment failures [95]. As an **example**, Hassan and Jalaludin [40] describes how a molding company implemented predictive maintenance algorithms to improve the operation of injection molding machines.

In the **near future**, predictive maintenance is advanced by LLMs. The factory is equipped with a network of sensors attached to every critical piece of machinery, collecting data on temperature, vibration, sound, and other metrics. This data is transmitted in real-time to the factory's central AI system. The system analyzes the incoming data. It detects subtle anomalies that might escape traditional monitoring systems, being able to predict equipment failures days or even weeks in advance. For example, the predictive system might notice a slight increase in vibration in a motor, which typically precedes bearing failure. Based on this analysis, the LLM generates maintenance schedules and sends alerts to the maintenance team, specifying which parts need attention and when. The LLM system also provides detailed diagnostic reports that explain the reasons behind each maintenance recommendation.

Using LLMs, **smart factories** can achieve higher product **quality**, reduce defects, and maintain consistent standards. This can ultimately lead to increased customer satisfaction and competitive advantage. As an illustrative example, consider the example of a smart automotive factory. The plant is equipped with numerous sensors and cameras installed in assembly lines, robots, and quality control stations. These devices continuously collect data on various parameters. As car parts move along the assembly line, the LLM processes real-time data. For example, it analyzes images captured by cameras to detect tiny defects in car body panels, such as dents, scratches, or misalignments. An abnormal deviation in the placement of a car door or an unusual vibration pattern can cause an alert. This scenario illustrates the significant impact that LLMs can have on current manufacturing processes and their importance in modern industry.

We can also imagine a **long-term** scenario of a highly automated smart factory that uses **advanced LLMs** for **quality control**. Intelligent sensors embedded throughout the factory capture detailed data on structural integrity, material composition, and environmental factors. LLMs process this data in real-time, detecting microscopic anomalies. Autonomous quality control agents conduct on-the-fly inspections and communicate with LLMs for continuous monitoring and immediate action. The factory's self-healing systems adjust production processes autonomously when defects are detected.

Continuous learning and **training** are essential to keep the workforce adaptable and capable of using new technologies. LLMs revolutionize **training** and knowledge transfer in manufacturing by providing real-time guidance and support to workers. They can serve as virtual mentors, offering step-by-step instructions, answering questions, and troubleshooting issues on the spot. This is particularly valuable for new employees or those dealing with complex machinery and processes. By offering personalized learning experiences and adapting to individual needs, LLMs can accelerate the learning process and help maintain a highly skilled workforce. Wang et al. [97] envision intelligent machines becoming coworkers rather than technological equipment. This is also the idea of Lemoine et al. [72] and Brondi et al. [11].

In the **future**, a smart factory will use LLM technology even more effectively. New workers participate in personalized training modules created by the LLM. These modules are tailored to their skills, experience, and learning preferences for efficient learning. Augmented reality (AR) glasses provide interactive, real-time instructions that guide employees through tasks with detailed, context-sensitive tutorials. Workers also practice complex tasks in virtual reality (VR) environments, gaining confidence in a risk-free setting. The LLM facilitates knowledge sharing, incorporating experience workers' insights into the training program.

Optimizing production schedules and **supply chain** management is essential in manufacturing operations. With LLMs, factories and industrial settings can produce dynamic production schedules that balance production capacity, inventory levels, and customer demand [58]. Work in this field is still in its early stages, although the use of LLMs for the supply chain holds significant promise.

The use of LLMs in manufacturing is a rapidly growing area of research with many applications [97]. However, to understand the advances in this field, it is essential to explore the history of natural language processing (NLP) and its origins. In the next section, we discuss the developments in natural language processing leading to the onset of Large Language Models.

3 Evolution of Natural Language Processing

Cambria and White [13] define natural language processing (NLP) as the set of computational techniques for the automatic analysis and representation of human language. The field lies at the confluence of linguistics, computer science, and artificial intelligence. NLP includes several sub-disciplines, such as language translation, text classification, sentiment analysis, and speech recognition [80]. These disciplines

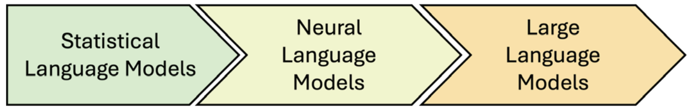


Fig. 3 Evolution in natural language processing

are central to many applications, including chat systems [103], recommender systems [112], information retrieval [7], and machine translation [19].

The origins of NLP can be traced to the 1950s [69], back to Alan Turing's pioneering work "Computing Machinery and Intelligence," [90] which introduced the Turing test. In the test, a human evaluator engages in a conversation with both a human and a machine through text-based communication, without knowing which is which. If the evaluator cannot reliably determine which participant is the machine based on the conversation alone, the machine is considered to have passed the Turing test. Turing's concept established the foundational principles for the field of NLP. A significant step towards passing the Turing test has emerged from the field of NLP with the advent of Large Language Models (LLMs). These models can generate human-like text, enabling them to engage in highly convincing conversations. In this section, we describe the evolution of NLP from its early days to the technology of LLM (see Fig. 3).

3.1 Statistical Language Models

In the early days of NLP, language models relied mainly on statistical language models, while machine learning was largely unexplored [43]. Claude Shannon [81] was one of the first to introduce the idea of using the statistical properties of text to model language. The most popular statistical language model, the n-grams, was introduced in his seminal work, the 1948 paper, "A Mathematical Theory of Communication." Shannon proposed n-grams to study the redundancy and predictability of English text. Following Shannon's initial work, the concept of n-grams was further developed and formalized in the field of computational linguistics and language modeling during the 1950s and 1960s [36].

The n-grams continues to be one of the most used statistical language models even today [1, 84]. This can be explained by its apparent simplicity and good performance [36]. The intuition behind the algorithm is that instead of using the entire history of words, we can instead rely only on the last few words. This means that you calculate the probability of the next word conditioned on its immediate $n - 1$ preceding words [66]. This basic idea has helped produce many advances that have improved the way we understand, recognize, and translate natural language.

The problem of n-grams models is that many word combinations do not appear in the training set. This is called the data sparcity problem [5]. The problem arises because the algorithm relies on word counts, which can result in certain unseen words never being considered for transcription and translation. For example, consider that you have seen the sentence "The dog has entered the bedroom" in the training corpus. The system cannot generalize to the sentence "The dog has entered the building," if the word "building" does not appear in combination with the previous words. One solution for this is to use smoothing techniques [20]. These methods adjust word distributions by increasing low probabilities and decreasing high probabilities.

3.2 Neural Language Models

Neural language models deal with data sparcity differently than the n-grams algorithm. Neural networks embed words in a continuous space where semantically similar words are close to each other. With these embedded vectors, the inference of probabilities is performed using the hidden layers of the network. In contrast to statistical language models, neural networks automatically learn features and capture long-range dependencies [3, 49].

Around the 1990s, recurrent neural networks (RNNs) [37] and their variants, such as long-short-term memory (LSTM) networks [41], began to show promise in the handling of sequential data by capturing dependencies in text. However, despite their success, these models still struggled with long-range dependencies and the need for manual feature engineering [24, 34, 110].

The introduction of the transformer architecture by Vaswani et al. [93] helped advance NLP in a spectacular way. Unlike RNNs, transformers can process entire sequences of text simultaneously, making them more efficient in capturing long-range dependencies [107, 111]. The key innovation is the self-attention mechanism, which allows the model to weigh the importance of different words in a sequence when making predictions [60].

A groundbreaking neural language model using the transformer architecture was introduced by Devlin et al. [25]. This model would start the era of large language models (LLMs). The model was named BERT (Bidirectional Encoder Representations from Transformers), and it was a Google invention. Importantly, it allowed the system to consider the context from both the left and the right of a word because of its bidirectional technology. BERT was pre-trained on a large corpus of text and fine-tuned for specific tasks. Its impact was significant, leading to subsequent developments in the ChatGPT series [38].

3.3 Large Language Models

A large language model (LLM) is a type of neural language model trained on vast amounts of data to understand, generate, and manipulate human language [17]. These models are typically based on neural networks such as transformers. The goal is to learn patterns, context, and relationships within text. These models perform a variety of tasks, such as translation, summarization, question answering, and conversational interaction.

GPT-1 [76] was introduced in 2018 based on generative pre-training (GPT). This was achieved by training on a very large dataset and using a transformer-based architecture that could generate text in an autoregressive manner. In the context of ChatGPT, "autoregressive" refers to the model that generates text one token at a time, where each token is predicted based on previously generated tokens [14]. In contrast, models like BERT do not generate text in an autoregressive manner. BERT is designed to predict missing words within a given context (masked language modeling) and is used to understand and classify text rather than sequentially generating it. Importantly, the model was trained on a diverse corpus of unlabeled text in a self-supervised learning way. GPT-2, released in 2019, further demonstrated the ability of an LLM to generate coherent and contextually relevant text [12].

GPT-3, introduced by OpenAI in 2020, expanded the concept of LLM even further with 175 billion parameters [30]. This complexity allowed it to perform a wide range of tasks just by understanding the context provided in the prompts. This capability is known as "few-shot" or "zero-shot" learning [75], in which a machine learning approach can make predictions about classes or tasks it has never seen during training.

Models like GPT-3 (and its successors) and BERT have found applications in various domains, from chatbots [79] and virtual assistants [87] to content creation [56] and coding assistance [74]. Their ability to understand and generate human-like text has transformed industries, making interactions more effective.

Both GPT-3 and GPT-3.5 models are not chat-optimized. This drawback is addressed by GPT-4 [70]. The system can also handle both text and image inputs. In addition to generating text with human-like fluency, GPT-4 models further pushed the results in many natural language processing tasks [45]. In addition to the ChatGPT family, the number of LLMs is growing [1] with other established families such as LLama and PaLM competing with ChatGPT. In the next section, we explain some of the technicalities of these systems.

4 Understanding Large Language Models

Large language models (LLMs) are artificial intelligence systems designed to understand, generate and manipulate human language [17, 48]. Formally, let $\mathcal{D}$ represent the training data set composed of text sequences and θ denote the LLM parameters. The model is trained to maximize the likelihood of the data:

$$\mathcal{L}(\theta) = \sum_{(x,y)\in\mathcal{D}} \log P(y \mid x; \theta)$$

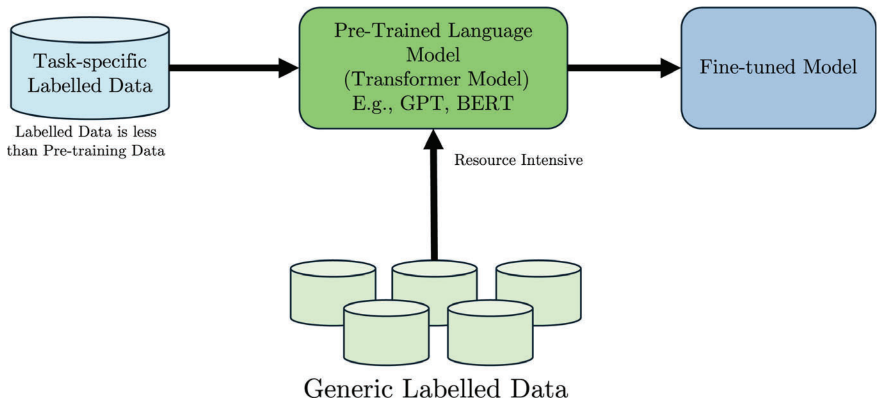


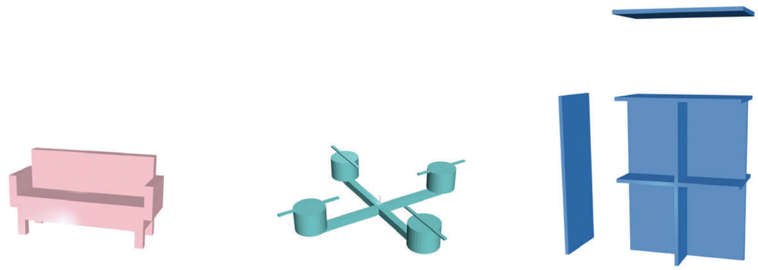
Fig. 4 Pretraining and post-pretraining of Large Language Models (LLMs)

where x is the input text sequence and y is the target text sequence. The probability $P(y \mid x; \theta)$ is computed using the transformer architecture, which incorporates a self-attention mechanism to model the dependenices between differnt parts of the input and target sequences.

The foundation of most LLMs is the transformer architecture, to which several authors have proposed improvements. Du et al. [27] explored the utility of model growth methods in LLM pre-training. Wu et al. [104] extended language models with the ability to memorize the internal representations of past input. Rawat et al. [78] modified the memory mechanisms. Geva et al. [33] explored the interpretability of these models. Wolf et al. [101] proposed a coding library with a variety of engineered transformer architectures.

LLMs undergo two stages of training: pretraining and post-pretraining [61] (see an illustration in Fig. 4). **Pretraining** occurs by making the model learn and extract patterns from large and diverse datasets. This stage allows the LLM to learn general language features and patterns. In the **post-pretraining** stage, the model is fine-tuned on specific datasets to specialize in particular tasks or domains. This process allows for greater adaptability and improved performance in a range of applications.

The generative capabilities of LLMs make them useful tools for various **applications**. For example, Wu et al. [102] implemented various conversational agents in the AutoGen framework. Here, LLMs are agents that converse with each other to perform tasks, such as a finance prediction or a mathematical task. Rangapur and Rangapur [77] compare different LLMs in their capacity as conversational agents. The authors analyzed the responses to various questions from ChatGPT, GPT-4, Gemini, Mixtral, and Claude. Liao et al. [59] discussed the role of proactiveness in conversational LLMs. Regarding automated content creation, several works utilize the capabilities of LLMs. Westerlund and Shcherbakov [100] discusses different possibilities in education, such as the design of workbooks to teach coding subjects. Moore et al. [67] reviews developments in automated content generation also for



(a) Meuble (b) Quad-copter Drone (c) Cabinet

Fig. 5 Assisted design with Large Language Model (LLM)

education. Ethape et al. [28] discusses how LLM generated content will change manufacturing and operation. Kim et al. [52] propose i-Dataquest for the manufacturing industry. This system is an information retrieval tool that uses data graphs.

In the following section, we investigate more extensively the potential of LLMs for smart manufacturing and Industry X.0.

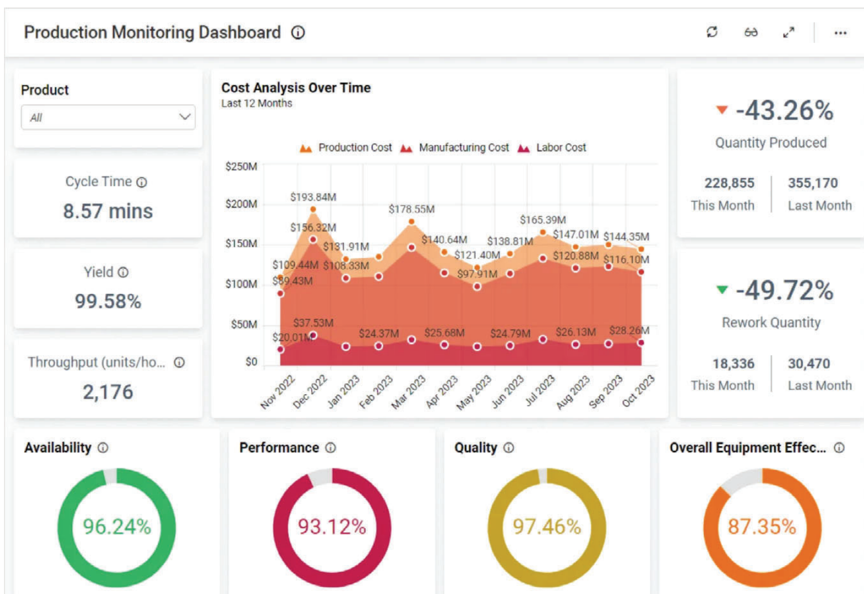
5 Applications of LLMs in Industry X.0

In this section, we describe three experiments using the Chat-GPT4 interface. The first experiment focused on design tasks, where the LLM generated multiple design options. The second experiment explored dashboard reporting, using Chat-GPT4 to interpret visual data and generate reports. The third experiment examined predictive maintenance, with Chat-GPT4 analyzing sensor data to predict equipment failures. Each experiment showcased the diverse applications of Chat-GPT4 in improving manufacturing processes.

5.1 Design Case Study

As an **application example**, we consider the work of Makatura et al. [63]. The authors studied and discussed the potential of LLMs for **design**. The authors assessed the abilities of LLMs in various design tasks, such as converting natural language prompts into designs or generating design spaces and variations. We followed some of their methods to produce the **examples** of Fig. 5. To create the images shown, we tasked ChatGPT-4 with generating JSCAD code for various objects. The instructions were minimal, such as "Generate me a JSCAD code of a couch meuble." Despite the prompt simplicity, the results shown in Fig. 5, were accurate. ChatGPT-4 effectively turned the instructions into detailed JSCAD code, producing precise designs.

However, there were challenges. Creating a coherent cabinet design was difficult. The generated design needed several iterations and refinements to meet the specifications. This issue highlighted areas for improvement in the system's understanding and generation capabilities. In contrast, the designs for the coach and the quad-copter drone were obtained without significant issues. These successes show the potential of ChatGPT-4 in generating complex design code with minimal input. The ability to produce accurate designs with limited instructions demonstrates the relative robustness of ChatGPT-4 in handling design tasks.



(a) Factory Dashboard

The "Production Monitoring Dashboard" provides a comprehensive overview of manufacturing metrics, including cycle time (8.57 mins), yield (99.58%), and throughput (2,176 units/hour). A cost analysis graph shows production, manufacturing, and labor costs over the past year. The dashboard highlights a significant decrease in quantity produced (-43.26%) and rework quantity (-49.72%) this month compared to last month. Key performance indicators include high availability (96.24%), performance (93.12%), and quality (97.46%), contributing to an overall equipment effectiveness (OEE) of 87.35%. This allows stakeholders to quickly assess production efficiency and identify areas for improvement.

(b) ChatGPT-4 Interpretation

Fig. 6 ChatGPT reads a productivity dashboard of a smart factory

5.2 Dashboard Case Study

As a practical example, consider Fig. 6, which shows a factory dashboard (Zerynth dashboard) and a Chat-GPT interpretation of the dashboard. Simply by feeding it a link to a Zerynth dashboard, the chatbot can read the data and graphs. It provides an overview of what is happening in real-time. The LLM interpretation is beneficial for an initial and superficial analysis. Importantly, the chatbot summarizes the factory's performance and which parameters are at their best. This immediate feedback helps factory managers make quick decisions, saving time compared to manually interpreting complex dashboards.

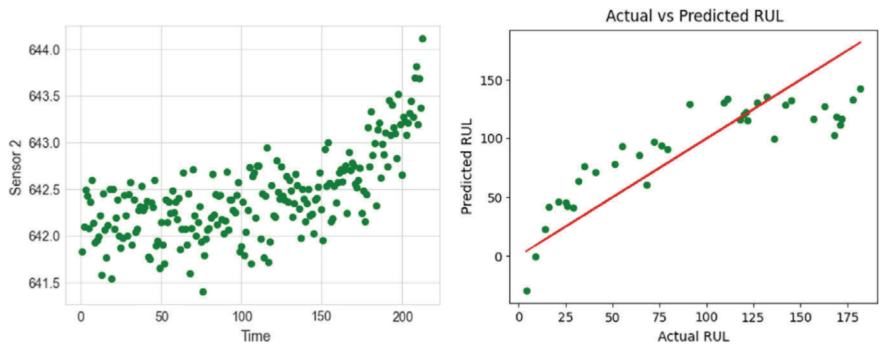
ChatGPT-4's answer effectively summarizes the existing information on the dashboard. It helps us understand key performance indicators clearly. Even if it does not produce new insights, its analysis aids operators who may be overwhelmed with data. It is also useful for generating the necessary documentation. This example highlights how ChatGPT-4 helps humans interpret data more easily. In general, using ChatGPT for dashboard interpretation has practical benefits. It is an asset for interpretability assessments in manufacturing environments.

5.3 Predictive Maintenance Case Study

To **analyze machinery data** and to address the task of **predictive maintenance**, LLMs can be explored in various ways [29]. LLMs can play an important role in tasks such as processing images and sensor data to identify defects during the manufacturing process. LLMs automate repetitive tasks and can analyze large data sets, optimize schedules, and improve **supply chain management**. As an **example**, we provided ChatGPT-4 with an image of a small sensory trajectory from C-MAPSS, a popular turbofan engine dataset (Fig. 7a). We then asked the LLM to produce predictions. Interestingly, the model was able to provide code with considerable accuracy. This result could be further analyzed and explored.

6 Challenges and Considerations

One of the technical challenges in leveraging data-driven methodologies, such as Large Language Models (LLMs), is ensuring high **data quality and availability**. For example, quality data helps in reducing biases. If the training data is representative and balanced, it minimizes the risk of the model having and amplifying societal biases. Agiza et al. [4] use Parameter-Efficient Fine-Tuning (PEFT) techniques to change the biases of LLMs. These techniques allow for the alignment of LLMs with targeted economic and political ideologies. This is especially important in manufacturing settings, since it is often necessary to align the assistant agents with the values and principles of the organization or company. Wang et al. [96] reviews



(a) C-MAPSS Trajectory

(b) Predictions with support of ChatGPT-4

Fig. 7 ChatGPT predicts failure time

bias (e.g., gender bias) in LLMs and also people's perceptions towards the present bias. They explore this in the context of the impersonation abilities of Chat-GPT and Vicuna-13B.

Data availability is often a challenge for LLMs in manufacturing, particularly in industries where data is siloed across different systems and departments. Kernan Freire et al. [47] introduced an LLM that extracted information in factory documentation and retrieved knowledge shared by manufacturing experts. Their factory study showed that LLMs can contribute to faster information retrieval and troubleshooting. However, the study also reported a preference for learning from a human expert when such an option is available. This finding may suggest the need for better affective computing even in industrial environments [73].

In manufacturing, data might be stored in **disparate formats** across various machines and processes, making it difficult to aggregate and analyze comprehensively. Overcoming these challenges may also require the help of LLMs. A data governance platform is proposed by Chen et al. [18]. Importantly, the platform integrates LLM, multi-agent system, and cloud native technology. Makatura et al. [62] also discusses the task of handling data formats in LLM manufacturing.

Another challenge of LLMs is the fact that the **computational resources** necessary to train machine learning models, particularly large language models (LLMs), are considerable. Training these models involves processing large amounts of data and performing iterative computations, which use significant computational power and memory. In addition, as LLM models grow in size and complexity, the computational resources required to train and deploy them scale accordingly. Marion et al. [64] worked on developing more effective training practices of LLMs using pruning techniques. Bi et al. [9] applied scaling laws to produce an LLM that had better performance than GPT-3.5.

Another topic of interest relates to the increase in the **lifecycle energy and carbon footprint** of LLM-powered intelligent chatbots [44]. Despite the ubiquity of

LLMs, the training, fine-tuning, and updating of such intelligent chatbots may result, in the future, in increased electricity consumption, resulting in carbon emissions. To address this problem, Jiang et al. [44] proposed a system-level approach with three strategies to optimize the management of Industry X.0 and mitigate the related footprints.

With increasing reliance on data-driven technologies, **data privacy and security** have become concerns. LLMs require robust security measures to protect against data breaches and unauthorized access [42, 91, 105]. Ensuring data privacy involves implementing encryption, access controls, and data storage to protect sensitive information. Van Bossuyt et al. [92] proposed ARCS-R. This methodology advocates the combination of safety and security in an overall assessment of resilience risk. Their case study involved an industrial inspection drone connected to the Internet with LLM and ML components.

Compliance with **data privacy regulations**, such as the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States, is essential. These regulations impose strict requirements on how personal data is collected, used, and shared, mandating transparency and user consent. Beyond data privacy, there are numerous industry-specific standards and regulations that organizations must adhere to when deploying AI technologies, such as LLMs. To date, there are no specific and general recommendations in manufacturing for AI and LLMs but these are soon expected, making it imperative for industrial organizations to prioritize compliance in their technological deployments.

As routine tasks become automated, the demand for skills such as problem solving, critical thinking, and **technical expertise** increases [62, 109]. Employees need to adapt to changing job requirements by developing proficiency in working with advanced technologies, understanding data analytics, and managing AI-driven systems. This shift necessitates a focus on upskilling and reskilling the workforce to ensure they can thrive in a technology-driven environment. Sheldon and Kwon [82] illustrates a case study of Samsung where the change in geographical location led to a reduction in the skilled workforce. To address the evolving demands of the job market, organizations must implement effective strategies for workforce training and upskilling. This can involve offering LLM-based training programs [71].

The deployment of Large Language Models (LLMs) in manufacturing introduces several ethical considerations. One primary concern is data privacy, as LLMs often require large volumes of data [62]. Ensuring that this data is collected, stored and utilized in compliance with privacy regulations is crucial to avoiding unauthorized access and misuse. In addition, there is the risk of bias in the models [50], which can lead to unfair or discriminatory outcomes in automated decision-making processes, potentially affecting hiring, human resources, and supply chain management.

Large language models (LLMs) development and deployment come with a set of challenges and considerations that must be addressed to achieve their full potential responsibly and ethically. In the next section, we review the future trends and developments of LLMs.

7 Future Trends and Developments

Current research in LLMs focuses on improving the **efficiency** of the Transformers architecture [94]. Techniques such as model distillation [89, 98], pruning [57], and quantization [106] aim to reduce the computational requirements of LLMs without losing performance. These advances are important for deploying LLMs in resource-constrained environments, such as edge devices and real-time applications. For example, Yin et al. [108] study the execution of LLM on the device. The goal here is also to preserve user privacy by running the agent on an edge device closer to the individual. A study on the environmental impact of LLM use in edge devices is published by Dhar et al. [26].

Interpretability is another area of ongoing research. Understanding how LLMs develop their predictions and decisions is essential for building trust and ensuring ethical AI [86]. Researchers are developing methods to make LLMs more transparent and explainable. For example, Singh et al. [85] reviews the emerging field of LLM interpretation (both interpreting LLMs and using LLMs for explanation). This is particularly important for applications in manufacturing [83].

The future of LLMs also lies in their ability to handle **multimodal data**—integrating text, images, audio, and other forms of data. This development is expected to unlock new applications in fields such as virtual reality and advanced human-computer interaction. Taylor et al. [88], for example, study the role of LLMs in a virtual reality setting (digital twin laboratory). Include a reference to Large Multi-modal Models here, Yang et al., The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision), 2023.

Multitask learning, in which a single model is trained to perform multiple tasks simultaneously, is another research direction. This approach can lead to more generalizable and accurate models, capable of transferring knowledge across different tasks and domains. For example, Chen et al. [21] introduce "tigerbot", a multi-language chatbot that achieves robust performance in different languages.

The **future of manufacturing with LLMs** will involve having smart factories where autonomous systems work together with human operators [46]. LLMs will allow machines to understand and respond to human commands in natural language [47]. These smart factories will take advantage of LLM analytics to optimize every aspect of production, from design, manufacturing, supply chain management, to quality assurance.

An example of integration of GPT and a real industrial setting is by Siemens. In this case study, the product is a car represented by the Simcenter Amesim vehicle dynamics model running on a Simcenter Prescan environment. The AI assistant (based on GPT), running in the background, continuously learns about the driver's preferences and the driver's reactions to changes in vehicle parameters. Although the focus of this project was car riding, the application of this technology could extend to a variety of other areas in design.

In conclusion, the **evolution of LLMs** and their capabilities may transform the manufacturing industry by improving automation, improving decision-making, and fostering innovation.

8 Conclusion

The current pace of technological advancement requires continuous innovation and adaptation in the manufacturing industry. Staying ahead in Industry X.0 requires a proactive approach to adopting and integrating technologies such as LLMs. Organizations must create a culture of innovation and exploration of emerging technologies.

Continuous learning and upskilling are essential to keep the workforce adaptable and capable of using new technologies. Providing access to training programs, workshops, and certification courses ensures that employees remain up-to-date. By investing in their workforce, organizations can build a proactive and innovative team capable of driving sustained growth and competitiveness.

The proactive adoption and exploration of LLMs are vital in manufacturing. Organizations should seek to understand the capabilities of LLMs and identify areas where they can drive improvements. Engaging with technology providers and leveraging their expertise can help organizations navigate the complexities of implementing LLMs. Building partnerships with technology vendors, consultants, and industry associations can provide valuable insight and resources for successful adoption. In addition, fostering a collaborative environment within the organization, where cross-functional teams work together to explore and implement LLM-based solutions, can accelerate the adoption process.

It is also important that organizations remain informed about the ongoing research and advances in LLMs. Keeping abreast of the latest developments can help organizations make informed decisions about technology investments. Participating in industry conferences, seminars, and forums provides opportunities to learn from experts and share experiences.

In conclusion, the evolution of LLMs and their capabilities marks a significant milestone in the journey of Industry X.0. Their transformative potential, driven by ongoing advancements and emerging use cases, promises to change manufacturing operations, driving efficiency, innovation, and agility. As LLMs continue to evolve, their integration into manufacturing processes will make the way for smarter, more efficient, and sustainable industrial operations, shaping the future of Industry X.0.

References

1. Mezzoudj F, Benyettou A (2018) An empirical study of statistical language models: n-gram language models vs. neural network language models
2. Achouch M, Dimitrova M, Ziane K, Sattarpanah KS, Dhouib R, Ibrahim H, Adda M (2022) On predictive maintenance in Industry 4.0: overview, models, and challenges. Appl Sci 12(16):8081
3. Adhikari S et al (2020) Nlp based machine learning approaches for text summarization. In: 2020 Fourth international conference on computing methodologies and communication (ICCMC), pp 535–538
4. Agiza A, Mostagir M, Reda S (2024) Analyzing the impact of data selection and fine-tuning on economic and political biases in LLMs. arXiv preprint arXiv:2404.08699
5. Allison B, Guthrie D, Guthrie L (2006) Another look at the data sparsity problem. In: Text, speech and dialogue: 9th international conference, TSD 2006, Brno, Czech Republic, September 11-15, 2006. Proceedings, vol 9, pp 327–334
6. Andarge T, Lichtenberg E (2020) Regulatory compliance under enforcement gaps. J Regul Econ 57(3):181–202
7. Ayoub MAK, Su Z, Li Q (2024) A case study of enhancing sparse retrieval using LLMs companion. In: Proceedings of the ACM on web conference, pp 1609–1615
8. Barker N, Jewitt C (2022) Future touch in industry: exploring sociotechnical imaginaries of tactile (tele) robots. Futures 136:102885
9. Bi X, Chen D, Chen G, Chen S, Dai D, Deng C, Ding H, Dong K, Du Q, Fu Z et al (2024) Deepseek llm: scaling open-source language models with longtermism. arXiv preprint arXiv:2401.02954
10. Brahmavar SB, Srinivasan A, Dash T, Krishnan SR, Vig L, Roy A, Aduri R (2024) Generating novel leads for drug discovery using LLMs with logical feedback. In: Proceedings of the AAAI conference on artificial intelligence, vol 38, no 1, pp 21–29
11. Brondi S, Pivetti M, Di Battista S, Sarrica M (2021) What do we expect from robots? Social representations, attitudes and evaluations of robots in daily life. Technol Soc 66:101663
12. Budzianowski P, Vulić I (2019) Hello, it's GPT-2—how can I help you? Towards the use of pretrained language models for task-oriented dialogue systems. arXiv preprint arXiv:1907.05774
13. Cambria E, White B (2014) Jumping NLP curves: a review of natural language processing research. IEEE Comput Intell Mag 9(2):48–57
14. Cao Y, Li S, Liu Y, Yan Z, Dai Y, Yu PS, Sun L (2023) A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to chatgpt. arXiv preprint arXiv:2303.04226
15. Carroni E, Delogu M, Pulina G (2023) Technology adoption and specialized labor. Int Econ 173:249–259
16. Casciani A, Bernardi ML, Cimitile M, Marrella A (2024) Conversational systems for AI-augmented business process management. In: International conference on research challenges in information science, pp 183–200
17. Chang Y, Wang X, Wang J, Wu Y, Yang, Zhu K, Chen H, Yi X, Wang C, Wang Y et al (2024) A survey on evaluation of large language models. ACM Trans Intell Syst Technol 15(3):1–45
18. Chen H, Lin X, Gong B (2024) Ansae: an Ai-Native Saas manufacturing and management platform based on Llm and multi-agent clustering system. Available at SSRN 4795529
19. Chen K, Chen B, Gao D, Dai H, Jiang W, Ning W, Yu S, Yang L, Cai X (2024) General2Specialized LLMs translation for E-commerce companion. In: Proceedings of the ACM on web conference, pp 670–673
20. Chen SF, Goodman J (1999) An empirical study of smoothing techniques for language modeling. Comput Speech Lang 13(4):359–394
21. Chen Y, Cai W, Wu L, Li X, Xin Z, Fu C (2023) Tigerbot: an open multilingual multitask LLM. arXiv preprint arXiv:2312.08688

22. Chopra R, Godbole AM, Sadvilkar N, Shah MB, Ghosh S, Gunning D (2020) The natural language processing workshop: confidently design and build your own NLP projects with this easy-to-understand practical guide
23. Colabianchi S, Tedeschi A, Costantino F (2023) Human-technology integration with industrial conversational agents: a conceptual architecture and a taxonomy for manufacturing. J Ind Inf Integr 35:100510
24. De Mulder W, Bethard S, Moens M-F (2015) A survey on the application of recurrent neural networks to statistical language modeling. Comput Speech Lang 30(1):61–98
25. Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
26. Dhar N, Deng B, Lo D, Wu X, Zhao L, Suo K (2024) An empirical analysis and resource footprint study of deploying large language models on edge devices. In: Proceedings of the 2024 ACM southeast conference, pp 69–76
27. Du W, Luo T, Qiu Z, Huang Z, Shen Y, Cheng R, Guo Y, Fu J (2024) Stacking your transformers: a closer look at model growth for efficient LLM pre-training. arXiv preprint arXiv:2405.15319
28. Ethape P, Kane R, Gadekar G, Chimane S (2023) Smart automation using LLM. Int Res J Innov Eng Technol 7(11):603
29. Design and operation of a GPT-powered smart manufacturing system (2024)
30. Floridi L, Chiriatti M (2020) GPT-3: its nature, scope, limits, and consequences. Minds Mach 30:681–694
31. Freire SK, Wang C, Niforatos E (2024) Chatbots in knowledge-intensive contexts: comparing intent and LLM-based systems. arXiv preprint arXiv:2402.04955
32. Galindo JA, Dominguez AJ, White J, Benavides D (2023) Large language models to generate meaningful feature model instances. In: Proceedings of the 27th ACM international systems and software product line conference, vol A, pp 15–26
33. Geva M, Caciularu A, Wang KR, Goldberg Y (2022) Transformer feed-forward layers build predictions by promoting concepts in the vocabulary space. arXiv preprint arXiv:2203.14680
34. Giles CL, Chen D, Sun G-Z, Chen H-H, Lee Y-C, Goudreau MW (1995) Constructive learning of recurrent neural networks: limitations of recurrent cascade correlation and a simple solution. IEEE Trans Neural Netw 6(4):829–836
35. Gkournelos C, Konstantinou C, Makris S (2024) An LLM-based approach for enabling seamless Human-Robot collaboration in assembly. CIRP Ann
36. Goodman JT (2001) A bit of progress in language modeling. Comput Speech Lang 15(4):403–434
37. Grossberg S (2013) Recurrent neural networks. Scholarpedia 8(2):1888
38. Hao Y, Dong L, Wei F, Xu K (2019) Visualizing and understanding the effectiveness of BERT. arXiv preprint arXiv:1908.05620
39. Hartman SM et al (2012) Protecting accelerator control systems in the face of sophisticated cyber attacks. Oak Ridge National Laboratory
40. Hassan MF, Jalaludin IM (2016) Application of why-why analysis to improve predictive maintenance strategy for injection molding machine. Inf Technol J 15(4):130–136
41. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural comput 9(8):1735–1780
42. Iqbal U, Kohno T, Roesner F (2023) LLM platform security: applying a systematic evaluation framework to OpenAI's ChatGPT plugins. arXiv preprint arXiv:2309.10254
43. Jiang K, Lu X (2020) Natural language processing and its applications in machine translation: a diachronic review. In: 2020 IEEE 3rd international conference of safe production and informatization (IICSPI), pp 210–214
44. Jiang P, Sonne C, Li W, You F, You S (2024) Preventing the immense increase in the life-cycle energy and carbon footprints of LLM-powered intelligent chatbots. Engineering
45. Kalyan KS (2024) A survey of GPT-3 family large language models including ChatGPT and GPT-4. Nat Lang Process J III 6:100048

46. Kernan FS, Foosherian M, Wang C, Niforatos E (2023) Harnessing large language models for cognitive assistants in factories. In: Proceedings of the 5th international conference on conversational user interfaces, pp 1–6
47. Kernan FS, Wang C, Foosherian M, Wellsandt S, Ruiz-Arenas S, Niforatos E (2024) Knowledge sharing in manufacturing using LLM-powered tools: user study and model benchmarking. Front Artif Intell 7:1293084
48. Khan JY, Uddin G (2022) Automatic code documentation generation using gpt-3. In: Proceedings of the 37th IEEE/ACM international conference on automated software engineering, pp 1–6
49. Khan W, Daud A, Nasir JA, Amjad T (2016) A survey on the state-of-the-art machine learning models in the context of NLP. Kuwait J Sci 43:4
50. Khowaja SA, Khuwaja P, Dev K, Wang W, Nkenyereye L (2024) Chatgpt needs spade (sustainability, privacy, digital divide, and ethics) evaluation: a review. Cognit Comput 1–23
51. Kim G, Baldi P, McAleer S (2024) Language models can solve computer tasks. Adv Neural Inf Process Syst 36
52. Kim L, Yahia E, Segonds F, Véron P, Mallet A (2021) i-Dataquest: a heterogeneous information retrieval tool using data graph for the manufacturing industry. Comput Ind 132:103527
53. Kimemia J, Gershwin SB (1985) Flow optimization in flexible manufacturing systems. Int J Prod Res 23(1):81–96
54. Komissarov PV (2021) Comprehensive assessment of the base of mathematical modelling of production business processes. Eur Sci e-J 8(14):7
55. Kühnle H (2017) Lecture notes in manufacturing systems design and manufacturing process organisation: selected chapters from factory operations, factory planning, manufacturing enterprise organisation & cyber physical production
56. Kumar P, Manikandan S, Kishore R (2024) AI-driven text generation: a novel GPT-based approach for automated content creation. In: 2024 2nd International conference on networking and communications (ICNWC), pp 1–6
57. Lagunas F, Charlaix E, Sanh V, Rush AM (2021) Block pruning for faster transformers arXiv preprint arXiv:2109.04838
58. Li B, Mellou K, Zhang B, Pathuri J, Menache I (2023) Large language models for supply chain optimization. arXiv preprint arXiv:2307.03875
59. Liao L, Yang GH, Shah C (2023) Proactive conversational agents in the post-chatgpt world. In: Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval, pp 3452–3455
60. Lin T, Wang Y, Liu X, Qiu X (2022) A survey of transformers. AI Open 3:111–132
61. Liu L, Liu X, Gao J, Chen W, Han J (2020) Understanding the difficulty of training transformers. arXiv preprint arXiv:2004.08249
62. Makatura L, Foshey M, Wang B Hähnlein F, Ma P, Deng B, Tjandrasuwita M, Spielberg A, Owens CE, Chen PY et al (2024) How can large language models help humans in design and manufacturing? Part 2: synthesizing an end-to-end LLM-enabled design and manufacturing workflow. Harvard Data Sci Rev
63. Large Language Models for Design and Manufacturing (2024)
64. Marion M, Üstün A, Pozzobon L, Wang A, Fadaee M, Hooker S (2023) When less is more: investigating data pruning for pretraining LLMS at scale. arXiv preprint arXiv:2309.04564
65. Mayet F (2024) GAIA: a general AI assistant for intelligent accelerator operations. arXiv preprint arXiv:2405.01359
66. Minaee S, Mikolov T, Nikzad N, Chenaghlu M, Socher R, Amatriain X, Gao J (2024) Large language models: a survey II
67. Moore[1] S, Tong R, Singh A, Liu Z, Hu X, Lu Y, Liang J, Cao C, Khosravi H, Denny P et al (2023) Check for updates empowering education with LLMS—the next-gen interface and content generation artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky: 24th international conference, AIED, Tokyo, Japan, July 3–7, 2023. Proceedings, p 32

68. Morales MG, del Cerro VF (2024) Augmented reality in Industry 4.0 assistance and training areas: a systematic literature review and bibliometric analysis. Electronics 13(6):1147
69. Noguer AM (2024) Key milestones in natural language processing (NLP) 1950-2024
70. OpenAI R et al (2023) GPT-4 technical report. arXiv:2303.08774
71. Orenstrakh MS, Karnalim O, Suarez CA, Liut M (2023) Detecting LLM-generated text in computing education: a comparative study for chatgpt cases. arXiv preprint arXiv:2307.07411
72. Pacaux-Lemoine M-P, Trentesaux D, Rey GZ, Millot P (2017) Designing intelligent manufacturing systems through human-machine cooperation principles: a human-centered approach. Comput Ind Eng 111:581–595
73. Picard RW (2000) Affective computing
74. Poldrack RA, Lu T, Beguš G (2023) AI-assisted coding: experiments with GPT-4. arXiv preprint arXiv:2304.13187
75. Pourpanah F, Abdar M, Luo Y, Zhou X, Wang R, Lim CP, Wang X-Z, Wu QMJ (2022) A review of generalized zero-shot learning methods. IEEE Trans Pattern Anal Mach Intell 45(4):4051–4070
76. Improving language understanding by generative pre-training. (2018)
77. Rangapur A, Rangapur A (2024) The battle of LLMs: a comparative study in conversational QA tasks. arXiv preprint arXiv:2405.18344
78. Rawat AS, Zhu C, Li D, Yu F, Zaheer M, Kumar S, Bhojanapalli S (2020) Modifying memories in transformer models. In: International conference on machine learning (ICML). 2021
79. Rivas P, Zhao L (2023) Marketing with chatgpt: navigating the ethical terrain of gpt-based chatbot technology. AI 4(2):375–384
80. Schopf T, Arabi K, Matthes F (2023) Exploring the landscape of natural language processing research. arXiv preprint arXiv:2307.10652
81. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423
82. Sheldon P, Kwon S-H (2023) Samsung in Vietnam: FDI, business-government relations, industrial parks, and skills shortages. Econ Labour Relat Rev 34(1):66–85
83. Shi F, Zhang Y, Qu C, Fan C, Chu J, Jin L, Liu S (2024) Leveraging the power of large language models to drive progress in the manufacturing industry. In: 9th International conference on financial innovation and economic development (ICFIED 2024), pp 125–133
84. Siivola V, Pellom BL (2005) Growing an n-gram language model. Interspeech, pp 1309–1312
85. Singh C, Askari A, Caruana R, Gao J (2022) Augmenting interpretable models with LLMS during training. arXiv preprint arXiv:2209.11799
86. Singh C, Inala JP, Galley M, Caruana R, Gao J (2024) Rethinking interpretability in the era of large language models. arXiv preprint arXiv:2402.01761
87. Taneja K, Maiti P, Kakar S, Guruprasad P, Rao S, Goel AK, Jill W (2024) A virtual teaching assistant powered by ChatGPT. arXiv preprint arXiv:2405.11070
88. The rise of the AI scientist: unleashing the potential of Chat-GPT powered avatars in virtual reality digital-twin laboratories (2023)
89. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H (2021) Training data-efficient image transformers & distillation through attention. In: International conference on machine learning, pp 10347–10357
90. Turing AM (1950) Computing machinery and intelligence. Creat Comput 6(1):44–53
91. Ullah S, Han M, Pujar S, Pearce H, Coskun A, Stringhini G (2024) LLMs cannot reliably identify and reason about security vulnerabilities (yet?): a comprehensive evaluation, framework, and benchmarks. In: 2024 IEEE symposium on security and privacy (SP), pp 199–199
92. Van Bossuyt DL, Papakonstantinou N, Hale B, Arlitt R, Palatheerdham SR (2024) ARCS-R: mission critical combined reliability and cybersecurity systems engineering analysis. In: 2024 Annual reliability and maintainability symposium (RAMS), pp 1–8
93. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need Advances in neural information processing systems. 30
94. Wang C, Li M, Smola AJ (2019) Language models with transformers. arXiv preprint arXiv:1904.09408

95. Wang K (2016) Intelligent predictive maintenance (IPdM) system—Industry 4.0 scenario. WIT Trans Eng Sci 113:259–268
96. Wang L, Song M, Rezapour R, Kwon BC, Huh-Yoo J (2023) People's perceptions toward bias and related concepts in large language models: a systematic review. arXiv preprint arXiv:2309.14504
97. Wang T, Zheng P, Li S, Wang L (2024) Multimodal human-robot interaction for human-centric smart manufacturing: a survey. Adv Intell Syst 6(3):2300359
98. Wang W, Wei F, Dong L, Bao H, Yang N, Zhou M (2020) Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Adv Neural Inf Process Syst 33:5776–5788
99. Wang X, Anwer N, Dai Y, Liu A (2023) ChatGPT for design, manufacturing, and education. Procedia CIRP 119:7–14
100. LLM integration in workbook design for teaching coding subjects. 2024
101. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M et al (2019) Huggingface's transformers: state-of-the-art natural language processing. arXiv preprint arXiv:1910.03771
102. Wu Q, Bansal G, Zhang J, Wu Y, Zhang S, Zhu E, Li B, Jiang L, Zhang X, Wang C (2023) Autogen: enabling next-gen LLM applications via multi-agent conversation framework. arXiv preprint arXiv:2308.08155
103. Wu T, He S, Liu J, Sun S, Liu K, Han Q-L, Tang Y (2023) A brief overview of ChatGPT: the history, status quo and potential future development IEEE/CAA. J Autom Sin 10(5):1122–1136
104. Wu Y, Rabe MN, Hutchins DL, Szegedy C (2022) Memorizing transformers. arXiv preprint arXiv:2203.08913
105. Yao Y, Duan J, Xu K, Cai Y, Sun Z, Zhang Y (2024) A survey on large language model (llm) security and privacy: the good, the bad, and the ugly. High-Confid Comput 100211
106. Yao Z, Yazdani AR, Zhang M, Wu X, Li C, He Y (2022) Zeroquant: efficient and affordable post-training quantization for large-scale transformers. Adv Neural Inf Process Syst 35:27168–27183
107. Yates A, Nogueira R, Lin J (2021) Pretrained transformers for text ranking: BERT and beyond. In: Proceedings of the 14th ACM international conference on web search and data mining, pp 1154–1156
108. Yin W, Xu M, Li Y, Liu X (2024) LLM as a system service on mobile devices. arXiv preprint arXiv:2403.11805
109. Youndt MA, Snell SA, Dean JW Jr, Lepak DP (1996) Human resource management, manufacturing strategy, and firm performance. Acad Manag J 39(4):836–866
110. Zargar S (2021) Introduction to sequence learning models: RNN. GRU Department of Mechanical and Aerospace Engineering, North Carolina State University, LSTM
111. Zhang X, Su Y, Tripathi S, Tu Z (2022) Text spotting transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9519–9528
112. Zhao Z, Fan W, Li J, Liu Y, Mei X, Wang Y, Wen Z, Wang F, Zhao X, Tang J et al (2024) Recommender systems in the era of large language models (LLMs). IEEE Trans Knowl Data Eng
113. Zonta T, Da Costa CA, da Rosa RR, de Lima MJ, da Trindade ES, Li GP (2020) Predictive maintenance in the Industry 4.0: a systematic literature review. Comput Ind Eng 150:106889