Journal of Manufacturing Systems 82 (2025) 420–448

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Review

Industrial Foundation Models (IFMs) for intelligent manufacturing: A systematic review

Shuxuan Zhao [a,1], Sichao Liu [b,e,1], Yishuo Jiang [d,1], Bo Zhao [c], Youlong Lv [c], Jie Zhang [c], Lihui Wang [b], Ray Y. Zhong [a,*]

[a] Department of Data and System Engineering, The University of Hong Kong, Hong Kong, China
[b] Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
[c] Institute of Artificial Intelligence, Donghua University, Shanghai, China
[d] Department of Systems Engineering, Cornell University, Ithaca, USA
[e] Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

ARTICLE INFO

Keywords:
Intelligent manufacturing
Industrial Foundation Models (IFMs)
Large Foundation Models (LFMs)

ABSTRACT

The remarkable success of Large Foundation Models (LFMs) has demonstrated their tremendous potential for manufacturing and sparked significant interest in the exploration of Industrial Foundation Models (IFMs). This study provides a comprehensive review of the current state of IFMs and their applications in intelligent manufacturing. It conducts an in-depth analysis from three perspectives, including data level, model level, and application level. The definition and framework of IFMs are discussed with a comparison to LFMs across these three perspectives. In addition, this paper provides a brief overview of the advancements in IFMs development across different countries, institutions, and regions. It explores the current application of IFMs, including Industrial Domain Models and Industrial Task Models, which are specifically designed for various industrial domains and tasks. Furthermore, key technologies critical to the training of IFMs are explored, such as data pre-processing, model fine-tuning, prompt engineering, and retrieval-augmented generation. This paper also highlights the essential capabilities of IFMs and their typical applications throughout the manufacturing lifecycle. Finally, it discusses the current challenges and outlines potential future research directions. This study aims to inspire new ideas for advancing IFMs and accelerating the evolution of intelligent manufacturing.

1. Introduction

Intelligent manufacturing [1,2] is a broad field of manufacturing that integrates automation [3], artificial intelligence (AI) [4], and advanced manufacturing technologies [5] to optimise the manufacturing process. It empowers manufacturing enterprises to fulfil increasingly customised product demands with shorter lead-time [6] and higher quality [7]. In recent years, advancements in Cyber–Physical Systems (CPS) [8], Internet of Things (IoT) [1], and Digital Twin (DT) [9–11] have enabled the industrial data collection and integration across the entire manufacturing workflows, achieving real-time synchronisation of physical processes and information flows [12]. Moreover, the in-depth exploration of industrial AI [13] and big data analytics (BDA) technologies [14] has provided innovative pathways to address challenges in diverse industrial scenarios.

Over the past few decades, the development of intelligent manufacturing technologies has been significantly propelled by breakthroughs in AI technologies [15], as shown in Fig. 1. The term "artificial intelligence" was first introduced at the Dartmouth Conference in 1956, with its envisioned steps encompassing search, pattern recognition, learning, planning, and induction [16]. Furthermore, the integration of AI technologies and manufacturing has undergone extensive exploration. In 1961, the first digitally controlled programmable robot was deployed on an assembly line, marking a milestone in replacing human labour with automated systems.

Subsequently, with the development of Symbolic AI, knowledge-driven expert systems [17] have been investigated to establish industrial knowledge bases and replicate the decision-making process of experts [18]. These systems were designed to offer accurate solutions to customers and facilitate specific manufacturing tasks. However, they faced persistent challenges in acquiring comprehensive domain-specific knowledge [19].

* Corresponding author.
E-mail address: zhongzry@hku.hk (R.Y. Zhong).
[1] The authors contributed equally to this work.

https://doi.org/10.1016/j.jmsy.2025.06.011
Received 31 January 2025; Received in revised form 28 May 2025; Accepted 8 June 2025
Available online 8 July 2025

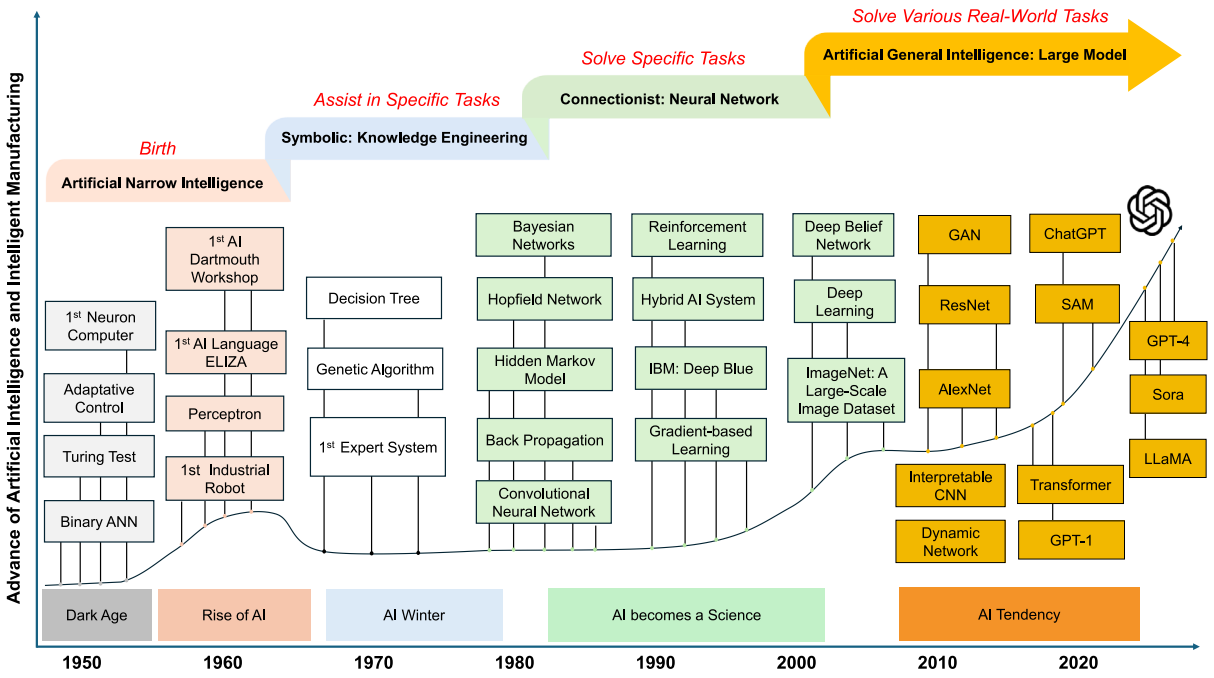

Fig. 1. The development of intelligent manufacturing and artificial intelligence [15].

Furthermore, breakthroughs in deep learning methods have positioned connectionist AI [20] as the core research focus, greatly accelerating the rapid development of intelligent manufacturing. Deep learning techniques leverage data-driven methods to train task-oriented models [21] using specific industrial datasets and neural networks such as Convolutional Neural Network (CNN) [22] and Recurrent Neural Network (RNN) [23]. By learning from industrial data, these models acquire powerful perception and decision-making abilities, enabling them to independently solve specific industrial tasks including predictive maintenance [24], fault diagnosis [25,26], quality control [27], and human-machine collaboration [28].

Although deep learning approaches have greatly improved manufacturing intelligence, their limited generalisation ability [29] still necessitates customised model training for specific tasks. Recently, with the emergence of Transformers [30], Large Foundation Models (LFMs) such as ChatGPT [31], Sora [32], SAM [33], and Claude [34] have achieved remarkable success in natural language processing (NLP), computer vision (CV), and data generation tasks. This has shifted deep learning approaches from single-task, single-modal, and limited data to a new paradigm encompassing various tasks, multimodality, and generalisation to large datasets [35]. The success of LFMs also demonstrates strong generalisation capabilities across various domains, showing significant potential for applications in manufacturing. Some recent studies [36] have demonstrated that Industrial-GPT can be integrated with advanced manufacturing technologies such as digital twin and knowledge graph to design the autonomous intelligent manufacturing system. With their exceptional performance in multi-level autonomous perception, cross-domain cognition, as well as event-driven collaborative decision-making, LFMs are regarded as crucial components of intelligent manufacturing systems. As a result, research on Industrial Foundation Models (IFMs) has attracted wide attention from both academia and industry [37].

However, when developing IFMs capable of addressing specific tasks in real-world industrial scenarios, several challenges remain to be solved. These challenges can be systematically categorised across three key dimensions: data level, model level, and application level.

Data level challenges: Industrial data exhibits the typical 3 V characteristics as variety, volume, and velocity [38]. Variety is reflected in the multi-source and heterogeneous industrial data, including sensor readings, video streams, text, and images. This requires IFMs to handle multimodal data and merge it into a unified representation, involving complex tasks such as multimodal feature extraction and fusion [39]. Volume refers to the massive and continuously growing size of industrial data, much of which consists of redundancy or low-density information. As a result, IFMs must precisely identify key information from vast amounts of data to enable intelligent perception in complex industrial scenarios [40]. Velocity implies high-frequency data generation and the need for real-time decision-making, requiring IFMs to process incoming data swiftly to meet real-time requirements.

Model level challenges: Industrial scenarios impose stringent requirements on models regarding real-time performance, robustness, reliability, and explainability [41]. Delayed responses or erroneous outputs from models can lead to direct economic losses or safety incidents [42]. Therefore, IFMs must be capable of balancing accuracy and real-time performance while maintaining strong robustness against disturbances and uncertainties in dynamic industrial environments. Besides, the black-box nature of deep learning restricts the application of IFMs in industries with rigorous safety standards, such as aerospace and healthcare. IFMs must be trustworthy and explainable enough to provide clear and traceable decision-making. As a result, designing IFMs that can meet the specific requirements of industrial scenarios is indeed a significant challenge.

Application level challenges: Industries involve a wide variety of process flows, physical phenomena, and specialised terminology [43]. To effectively address industrial tasks, IFMs must not only analyse multimodal data but also dynamically learn domain-specific knowledge, such as physical laws and process regulations. Efficiently embedding such domain knowledge into IFMs presents a significant challenge [43]. Moreover, IFMs need to be integrated with manufacturing systems to handle specific industrial tasks [44]. Industrial systems are typical complex systems with different architectures, communication protocols and requirements. Achieving seamless and efficient integration of IFMs into such complex systems without compromising performance constitutes another significant challenge.

To bridge these gaps and explore the impacts of IFMs in manufacturing, this paper aims to offer a systematic overview of the current

research on IFMs. The contributions of this paper can be summarised as follows:

(1) This paper clearly introduces a comprehensive framework for IFMs in the context of intelligent manufacturing. It encompasses the infrastructure layer, model layer, and application layer. The infrastructure layer serves as the foundation for IFMs, incorporating industrial data, computational resources, industrial knowledge, and programming frameworks. The model layer leverages multimodal industrial data and deep learning methods to build pre-trained models and further refines them by integrating knowledge and data for specific domains or tasks. The application layer is designed to seamlessly integrate IFMs with manufacturing systems, enabling them to address complex industrial tasks efficiently.

(2) Key technologies of IFMs are analysed from the perspectives of the data level, model level, and application level. Data analytics techniques, including data acquisition, data cleaning, data labelling, and data pre-processing, are summarised to analyse heterogeneous and multimodal data. Model development technologies are explored to leverage industrial data and knowledge to construct pre-trained foundation and industrial domain models. Finally, application technologies, such as prompt engineering, retrieval-augmented generation, and agent engineering, are introduced to facilitate the integration of IFMs with manufacturing systems for industrial tasks execution.

(3) This paper provides a comprehensive review of current research, technological development, and advancements in IFMs. It highlights domain-specific applications, such as robotics IFMs, automotive IFMs, and steel IFMs, demonstrating their adaptability across diverse industrial sectors. Furthermore, an in-depth discussion is conducted on six core capabilities and their applications in specific tasks. Finally, key challenges faced by IFMs and the future research directions are outlined.

The rest of this paper is organised as follows. Section 2 describes the definition and architecture of IFMs. Section 3 systematically reviews the current development and progress of IFMs. Section 4 presents the technologies of IFMs from the perspective of data, model, and application. Several typical applications of IFMs in industrial scenarios are also discussed in Section 5. Key challenges of IFMs are concluded in Section 6. Finally, this study's conclusions and future directions are given in Section 7.

2. Definition of IFMs

The release of ChatGPT has sparked a significant wave of research on LFMs, indicating that AI is entering a new phase towards Artificial General Intelligence (AGI) [45,46]. Compared to previous networks such as CNN and long short-term memory (LSTM) networks, LFMs that utilise the Transformer architecture and self-attention mechanism are particularly good at handling long-distance dependencies and capturing global information. Their design not only facilitates multimodal data fusion but also supports the construction of larger and more expressive networks, ultimately leading to superior generalisation capabilities across diverse applications [47]. Before the emergence of LFMs, designing specialised models for specific industrial tasks and scenarios was commonplace. The success of LFMs has also given rise to IFMs, which are expected to bring a new paradigm of intelligent manufacturing characterised by "one foundation model for diverse industrial applications" [46].

IFMs denote deep learning models with self-attention mechanisms, large-scale parameters, and multimodal learning capabilities that target applications across the entire lifecycle of industrial production [48]. Compared with LFMs, IFMs share similar structures but incorporate more industrial data and domain-specific knowledge such as physical laws, process regulations, and technical terminology [49], thus enabling them to better address industrial tasks [50]. IFMs are directly applied in industrial scenarios to generate specialised industrial content, provide highly reliable and trustworthy outputs, support cross-domain industrial tasks, and realise self-adaptation across various industrial scenarios. With core capabilities including question answering, scene cognition, decision-making, terminal control, content generation, and scientific discovery [51], IFMs are capable of meeting the requirements of various domains in both discrete and process industries. Ultimately, IFMs are integrated with industrial systems, providing a new methodological technology for a wide range of industrial tasks including research & design, manufacturing, operation & management, and maintenance. Commonly, the framework of IFMs can be presented across the infrastructure level, model level, and application level, as shown in Fig. 2.

(a) Infrastructure Level

The infrastructure level refers to fundamental resources required for IFMs, comprising industrial data, computational resources, industrial knowledge and programming frameworks [52].

Industrial data encompasses multimodal data collected from industrial scenarios, such as sensor data, equipment operational data, production process data, and historical maintenance data. This data is presented in the form of images, video, audio, text, CAX and time-series data [53]. Industrial data serves as the basis for the training of IFMs.

Computational resources encompass the computational power and storage capacity necessary for the training and inference of IFMs across cloud, edge, and on-device environments. Training IFMs requires substantial computational resources, including high-performance computing (HPC) clusters, cloud infrastructure, and accelerators such as GPUs or TPUs. These resources are essential for ensuring the efficient training of IFMs.

Industrial knowledge comprises both general domain knowledge and specialised knowledge from enterprises, mainly covering industry standards, operational documents, machine operation principles, and maintenance experience [54]. Industrial knowledge can be integrated into IFMs through rules, knowledge graphs, or expert systems [55]. Industrial knowledge provides a logical foundation for the decision-making processes of IFMs.

Programming frameworks are tools and libraries used to build, train, and deploy IFMs, such as TensorFlow and PyTorch. These frameworks leverage distributed training and parallel computing strategies to optimise the utilisation of computational resources. Efficient programming frameworks can significantly reduce the training and deployment costs of LFMs.

(b) Model Level

The model level is the core of IFMs, which utilises deep learning algorithms to train IFMs for industrial scenarios. IFMs can be further categorised into Industrial Domain Models (IDMs) and Industrial Task Models (ITMs) as their application scenarios, embedded knowledge, and training datasets [49].

IFMs are trained on large-scale public multimodal datasets. This pre-training approach enables IFMs to develop general capabilities for processing multimodal data and achieve excellent problem-solving abilities for industrial tasks. Although IFMs often cannot be directly applied to industrial scenarios, they provide a critical foundation for adapting models to meet the requirements of specific industrial scenarios [56].

IDMs are developed for industries such as manufacturing, energy, and chemical engineering, with the primary objective of addressing universal tasks within these domains. They are continuously fine-tuned from pre-trained IFMs with public industrial data and knowledge encompassing industry-standard specifications, general process parameters, and open-source production logs [57]. Equipped with such industrial data and knowledge, IDMs possess a comprehensive understanding of the business logic and production processes of the respective industry, thereby providing standardised solutions. However, IDMs lack detailed knowledge about specific industrial tasks such as business needs, operational models, and process parameters because manufacturing enterprises often have confidential requirements, so their applications in specific industrial tasks still have limitations.

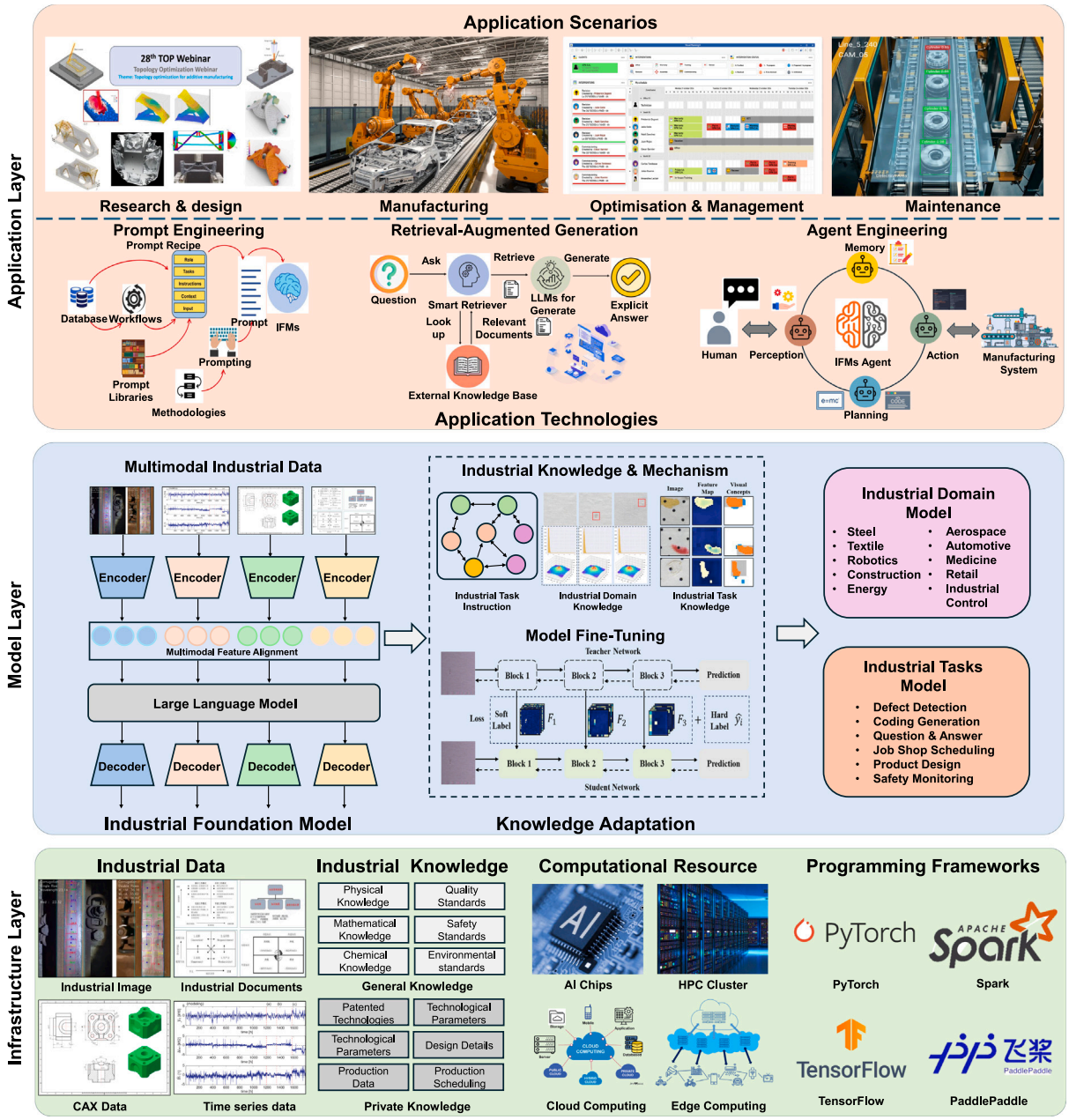

Fig. 2. The architecture of IFMs with infrastructure, model and application layers.

ITMs are purpose-built for specific industrial scenarios or tasks within individual enterprises, such as optimising customised assembly processes for a particular automotive manufacturer or regulating energy consumption in specific power plants. Typically derived from IDMs, ITMs undergo secondary fine-tuning or transfer learning using proprietary enterprise data, including exclusive process parameters, confidential production workflows, and customised business rules. By integrating specialised data and knowledge of the enterprise, ITMs are able to efficiently solve specific industrial tasks [58]. Besides, ITMs are also deeply adapted to enterprise-specific requirements of industrial tasks such as real-time performance, robustness, and accuracy.

(c) Application Level

The application level includes application technologies of IFMs and feasible application scenarios for IFMs. It is responsible for integrating IFMs into manufacturing systems and providing the necessary services [59].

Application technologies include prompt engineering, retrieval-augmented generation (RAG), agent engineering, and collaboration among IFMs. Prompt engineering aims to develop effective prompts to optimise the interaction between IFMs and users [60]. Prompt engineering provides guidance for user input commands, directing the system to execute specific tasks or generate specific content. It can bridge the gap between IFMs and industrial tasks, enabling IFMs to excel in various industrial scenarios. RAG involves developing online-updating external knowledge bases and information retrievers to enhance the performance of IFMs [61]. By incorporating the latest task-oriented knowledge, the inference of IFMs can become more traceable and interpretable, leading to reliable outputs. Agent engineering is responsible for the interactions between IFMs, humans, and manufacturing systems. The Agent enables manufacturing systems to achieve the closed-loop control and optimisation of "perception, analysis, decision-making and action" with the help of IFMs [62]. Besides, it can also fully take advantage of various IFMs to meet different industrial requirements such as accuracy, robustness, and real-time performance.

IFMs can be widely applied throughout the entire manufacturing lifecycle, including research & design, manufacturing, operation & management, and maintenance services [38]. IFMs encompass six core capabilities: question answering, scene understanding, process decision-making, terminal control, content generation, and scientific discovery [52]. These capabilities enable IFMs to be applied across various industrial scenarios. Question answering provides on-demand industrial literature retrieval, analysis, and Q&A services for industrial tasks [57]. Scene understanding capabilities can analyse complicated industrial environments, such as defect detection, fault analysis, and safety monitoring [22]. IFMs can also offer suggestions and make decisions based on knowledge and historical experience, such as production scheduling and emergency responses [63]. Besides, the terminal control capabilities enable IFMs to be applied in embodied intelligence and human–robot collaboration (HRC) [64,65]. Content generation capabilities can be used to generate coding, application, documentation, email, and CAX models [66]. Scientific discovery capabilities can help identify mechanisms between mechanical, electrical, hydraulic, thermal, pneumatic, and magnetic interactions within products, thereby revealing the physical and chemical principles for new product designs [67].

3. Development status of IFMs

This section provides a comprehensive overview of the development of IFMs by presenting current research publications on LFMs across various subjects from the Scopus and Google Scholar databases. Next, it discusses current IFMs launched by AI companies such as Google and Huawei. Furthermore, it introduces the current landscape of IDMs and ITMs that are applicable to specific industrial scenarios and tasks.

3.1. Research publications

This review examines research publications on LFMs, including the LLMs, Large Vision Models (LVMs), and Large Multimodal Models (LMMs) across different subjects. Overall, the research publication is introduced from four main perspectives: (a) publications by years, (b) publications by subjects, (c) publications by regions, (d) publications by affiliations, and (e) overview of the literature survey.

(a) Publications by Years

Fig. 3(a) illustrates the publication trends of LLMs from 2017 to 2024. From 2017 to 2021, research on LLMs experienced steady growth, with 2021 marking a pivotal turning point as academia and industry began prioritising the study of LLMs [68]. The number of LLM publications reached 104 in 2022, 1706 in 2023, and 4761 in 2024, respectively. It demonstrates a remarkably accelerated growth rate driven by the excellent performance of LLMs across a broad range of tasks.

Moreover, the release of GPT-3 in mid-2020 played a critical role in sparking the research interest in LLMs [69]. Starting in 2022, annual growth in publication counts more than doubled, underscoring a surge in motivation and investment in the field. This exponential increase can be attributed to the growth of large-scale datasets, significant advancements in computational resources, the widespread adoption of LLMs across various applications, and the public release of their source code.

(b) Publications by Subjects

Fig. 3(b) illustrates the distribution of LLMs research across different subjects. Computer science stands as the largest contributor to LLMs research, accounting for 37.5% of related publications. It primarily investigates core models and internal mechanisms of LLMs, including network architectures, optimisation algorithms, and training approaches. Novel attention mechanisms, parameter-efficient training techniques, and transfer learning approaches are designed to improve model performance and efficiency [70,71]. Besides, it also explores the integration of LLMs with other technologies (e.g., computer vision and reinforcement learning) to develop multimodal LLMs capable of handling more complex tasks.

Engineering ranks as the second-largest subject area with 11.9% contributions of publications, focusing on the applications of LLMs in industrial scenarios. It mainly aims to integrate LLMs into industrial systems to address real-world challenges. Compared to computer science, engineering emphasises leveraging LLMs to address specific industrial challenges and meet particular demands, such as real-time performance, energy efficiency, and robustness [72]. Moreover, LLMs are often customised for specific industrial tasks, such as intelligent product design and 2D-CAD drawings generation.

Mathematics contributes approximately 15% of LLMs' research publications, focusing on theoretical foundations and architectural optimisation. Researchers in this domain aim to develop mathematical frameworks to enhance LLMs' efficiency and interpretability, including optimisation algorithms (e.g., stochastic gradient descent variants, adaptive learning rate schedules) and theoretical models (e.g., neural tangent kernels, Lipschitz continuity analysis). Mathematics provides deeper insights into the fundamental workings of LLMs. Social science, arts, physics, astronomy, and other subjects collectively constitute about 40% of the research publications.

(c) Publications by Regions

Fig. 3(c) illustrates the distribution of LLM publications by region, with prominent contributions from the United States, China, and Europe. The United States leads in publications driven by major technology companies and top academic institutions. Renowned universities such as Stanford, MIT, and Carnegie Mellon are at the forefront of advancing theoretical frameworks, model architectures, and ethical AI. Meanwhile, companies like Google, OpenAI, Microsoft, and Meta are driving groundbreaking research and applications in LLMs. Innovations such as Google's LaMDA [73], OpenAI's GPT series [69,74], Microsoft's Turing-NLG [75], and Meta's LLaMA [76] exemplify U.S.-led advancements that continue to shape the global AI landscape.

China ranks second in LLMs research, driven by substantial investments in artificial intelligence and a strong focus on technological advancements. Leading institutions such as Tsinghua University [77], Peking University, and the Chinese Academy of Sciences have played pivotal roles in LLMs research, emphasising model efficiency and scalability. Companies like Baidu, Alibaba, Tencent, and Huawei are also at the forefront of LLMs development. Notable models include Baidu's Ernie [78], Alibaba's M6 [79], Tencent's Hunyuan [80], and Huawei's Pangu [81], which span a wide range of applications from natural language understanding to industrial intelligence.

Europe is the third-largest contributor, with key countries including the United Kingdom, Germany, and France. Institutions like the University of Oxford, ETH Zurich, and Sorbonne University lead efforts in responsible AI practices and understanding the societal impacts of LLMs. For example, they focus on developing ethical frameworks to address bias in generative models and studying how LLMs impact labour markets and misinformation landscapes. This global distribution highlights the competitive and collaborative nature of LLMs research, with each region contributing unique strengths in advancing the field.

(d) Publications by Affiliations

Fig. 3(d) illustrates affiliations of LLM publications. The major contributions come from China, the United States, and Europe, with diverse leadership across academic, industrial, and governmental sectors. The Chinese Academy of Sciences (1082) and Tsinghua University (950) top the ranking list, reflecting the substantial investment of China in AI research. Other prominent Chinese institutions, such as Peking University (681), the University of Chinese Academy of Sciences (653), and Zhejiang University (523), also play pivotal roles. In the United States, Carnegie Mellon University (884), Stanford University (751), and MIT (495) are at the forefront of LLMs research, often driven by partnerships with leading players and supported by significant

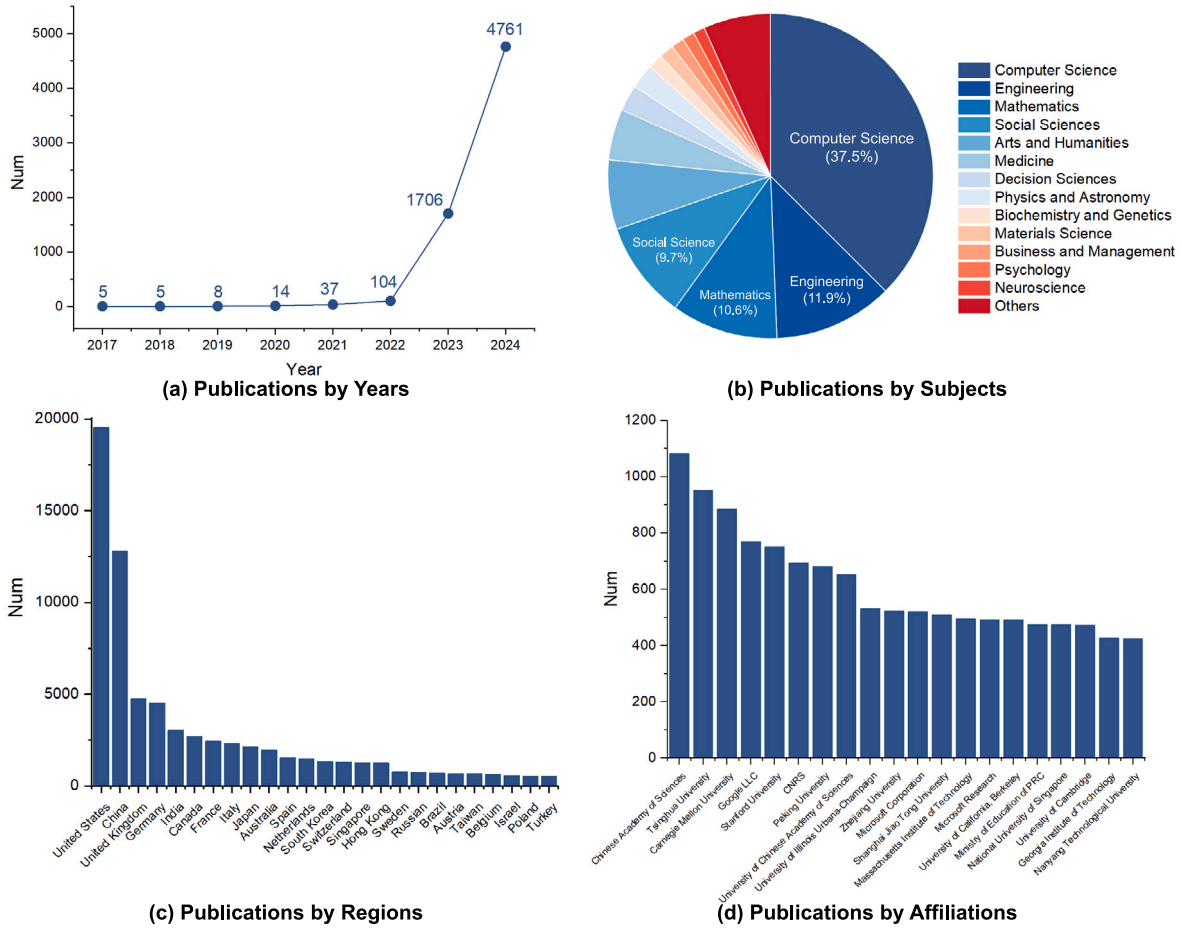

Fig. 3. Statistics from Scopus database (Search keywords: "LLM & LVM & LMM"; Date: 27 December 2024).

government funding. Google LLC (768), Microsoft Corporation (521), and Microsoft Research (492) represent the critical role of the private sector in bridging the gaps between theoretical innovations and commercial applications. In Europe, institutions like Centre National de la Recherche Scientifique (693) and the University of Cambridge (473) are noteworthy contributors, focusing on ethical AI practices, regulatory compliance, and fundamental research that addresses key challenges in transparency and fairness. The National University of Singapore (475) and Nanyang Technological University (425) reflect substantial contributions from Asia outside China, emphasising the importance of multicultural perspectives in LLMs development.

(e) Overview of Literature Survey

Research on LFMs and IFMs has also been extensively conducted, with several review papers already published. To clarify the distinction of this study, similar review studies on LFMs and IFMs are listed in Table 1.

Papers 1, 2, and 3 focus on LFMs for NLP and multimodal tasks. Paper 1 focuses on the foundational concepts, history, and recent advancements of LLMs. It extensively covers the development of model architectures, training strategies, fine-tuning techniques, evaluation metrics, and applications. Paper 2 focuses on pre-trained foundation models like BERT and GPT and their evolution across NLP, computer vision, and graph learning. It emphasises the cross-domain abilities of pre-trained foundation models and their applications in multimodal data processing. Paper 3 highlights the efficiency of LMMs, focusing on lightweight models and computational optimisation. It discusses how LMMs can be optimised for tasks like visual question answering and image understanding.

Papers 4 and 5 focus on the application of IFMs in industrial and manufacturing settings, highlighting specific challenges such as high trustworthiness, real-time inference, and multi-process coordination. Paper 4 explores the challenges and opportunities of applying IFMs in industrial applications with a focus on intelligent manufacturing. It proposes a new architecture and highlights the importance of real-time inference and high-accuracy requirements. Paper 5 discusses the application of IFMs in intelligent manufacturing. It contrasts the limitations of deep learning in manufacturing with the capabilities of IFMs, emphasising their powerful generalisation and ability to address complex tasks like fault diagnosis and quality control.

This paper provides a detailed discussion of IFMs from three perspectives: data level, model level, and application level. It introduces a comprehensive overview of data processing, model development, and model application of IFMs for intelligent manufacturing. First, the definition of IFMs is introduced with a comparison to ILMs. Then, a detailed framework of IFMs containing infrastructure, model, and application layer is designed. Second, this paper comprehensively reviews key technologies, including data analytics, development, and application technologies for IFMs. Third, it highlights the core functions and typical applications of IFMs. Finally, the challenges and future directions of IFMs for intelligent manufacturing are summarised.

3.2. Current IFMs

This section provides an outline of the evolutionary process of LFMs and IFMs in recent years. Each generation has iterated on the previous one and introduced new functions. The timeline and milestones of the existing LFMs and IFMs are shown in Fig. 4.

• 2019–2020: Early large models including BERT [85], mT5 [86], and GPT-2 [51] utilise the Transformer and self-attention mechanism

Table 1
Review papers on LFMs and IFMs.

No.	Title	Year
1	A Survey of Large Language Models [68]	2024
2	A comprehensive survey on pre-trained foundation models: a history from BERT to ChatGPT [82]	2024
3	Efficient Multimodal Large Language Models: A Survey [83]	2024
4	Industrial foundation model: Architecture, key technologies, and typical applications [84]	2024
5	Large Scale Foundation Models for intelligent Manufacturing Applications: A Survey [37]	2025

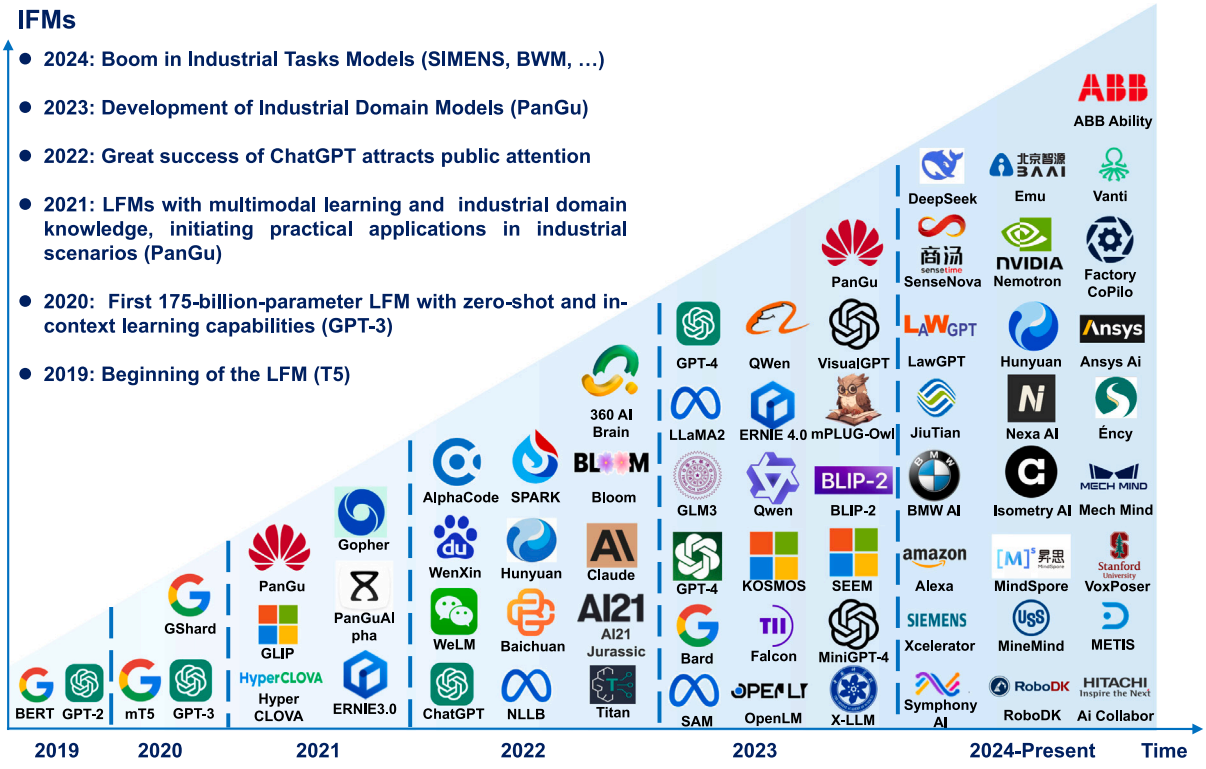

Fig. 4. A timeline of existing LFMs and IFMs.

to construct LFMs, achieving great performance in NLP tasks. Although mainly concentrated on NLP tasks and not directly applicable to diverse industrial tasks (e.g., visual detection, fault diagnosis, scheduling), the substantial breakthroughs of these models in generalisation capabilities greatly promoted the development of LFMs. A pivotal milestone emerged when OpenAI introduced GPT-3 [69], a 175-billion-parameter model capable of addressing varied tasks through "zero-shot learning" and "in-context learning" without task-specific fine-tuning. This marked the arrival of the large model era. While GPT-3 was primarily designed for general-purpose use, its few-shot learning and generalisation capabilities laid the foundation for applications in industrial scenarios.

• 2021: By approximately 2021, research interest increasingly shifted towards multimodal and knowledge-enhanced models like ERNIE 3.0 [78], Pangu-Alpha [81], and Inspur Yuan 1.0 [87], which were designed to integrate domain-specific knowledge into LFMs. Compared to previous models, they can not only handle various tasks but also enhance their performance by leveraging industrial domain knowledge, thus evolving into IFMs. By embedding domain knowledge, these IFMs can enable more robust applications in industrial tasks such as information processing, content generation, and text analysis. The Pangu-Alpha employs a three-layer architecture— "foundation model + industry model + scenario model" to incorporate industrial mechanism knowledge into model training. It has been applied in industrial scenarios to address practical tasks such as multimedia retrieval, fault diagnosis, visual defect detection, and software development.

• 2022: In 2022, ChatGPT achieved enormous commercial success with its exceptional conversational interaction capabilities, bringing LFMs into the public spotlight and significantly accelerating the development of LFMs. Both academia and industry have actively pursued LFMs capable of achieving multimodal information integration and multi-task decision-making [88,89] within unified frameworks, aiming to achieve "one model for multiple tasks" [90,91]. The emergence of multimodal LFMs, including BLOOM [92], iFLYTEK Spark [93], AlphaCode [94], and Titan, introduced a paradigm shift from predominantly language-focused processing to handling diverse modalities (texts, images, videos, codes) and multiple tasks. Due to their great capabilities for multimodal data analysis, these IFMs are particularly well-suited to industrial scenarios with heterogeneous task requirements. These state-of-the-art pursuits are jointly propelling LFMs from the realm of laboratory research into large-scale commercial utilisation and industrial applications.

• 2023: Entering 2023 and beyond, the development trend of LFMs is moving towards more universal cross-modal intelligence and domain-driven adaptation. General-purpose intelligence models such as GPT-4 [74], LLama2 [76], Qwen 2.0 [77], mPLUG-Owl [95], ImageBind [96], VisualGPT [97], SEEM, BLIP-2 [98], ERNIE 4.0 [78], GLM3, and KOSMOS [99] represent a new generation of models equipped with advanced cross-modal reasoning capabilities and extensive adaptability across domains. These models possess a larger number of parameters and more powerful multimodal feature extraction, enabling them to tackle more complex problems. Moreover, Huawei has upgraded its Pangu 3.0 with a focus on seven core industries, including energy,

finance, and manufacturing. Concerns about robustness, explainability, safety, compliance, and data privacy have gained more attention, leading researchers to refine models to meet the stringent standards required in industrial implementation [100,101].

• 2024-present: From 2024 to the present, with the maturation of ILMs technologies, their development costs have significantly decreased. Research on IFMs for industrial domains or tasks has also undergone rapid advancement, with models such as CivilGPT [51], Haier Home-GPT, ManipLLM [102], and Siemens Industrial Copilot being developed to solve specific tasks. Compared with previous IFMs, these models are primarily created by manufacturing enterprises, including Siemens, Haier, and Xiaomi. Leveraging their rich industrial data, clear understanding of industrial tasks, and the existing high-performing LFMs, these manufacturing enterprises intend to develop IFMs that can be directly deployed in industrial scenarios. Furthermore, these IFMs are engineered to be integrated with manufacturing systems in real-world industrial scenarios, so as to effectively solve practical industrial tasks.

With six core capabilities, including question answering, scene understanding, process decision-making, terminal control, content generation, and scientific discovery. IFMs have been widely applied throughout the entire manufacturing process to improve the efficiency of specific tasks such as product design, HRC assembly, and manufacturing process optimisation. The IFMs can be utilised to facilitate the product design, including the design formation, design interactions, and design optimisation [103]. For instance, the LLM-augmented multimodal collaborative design framework [104] is designed to provide professional design prompts and generate precise visual schemes. Besides, the mixed reality is used to form an interactive and immersive environment for users to participate in the design process. By integrating these technologies, it is able to form a unified cognition and optimise the traditional collaborative product design process. To achieve more effective HRC assembly, RoboFlamingo [105] combines vision-language models with imitation learning. By leveraging its vision understanding and content generation capabilities, this model can perceive the positions of parts and generate robot trajectories, enabling robots to autonomously carry out industrial assembly tasks. The assembly IFM [106] is also designed to achieve natural language understanding and behaviour-based control for robots. It is able to facilitate intuitive interactions and greatly improve the assembly efficiency in HRC. IFMs are also widely applied for the optimisation of the manufacturing process. IFMs [107] are employed to enable natural language queries and achieve flexible data visualisation. This allows production personnel to effectively interact with the process data generated by the manufacturing system and optimise the manufacturing process. In order to better understand the knowledge of the manufacturing process, knowledge graphs are also utilised to embed multimodal domain knowledge into IFMs. This aims to achieve a unified representation of process knowledge and to leverage the excellent perception and decision-making capabilities of IFMs to analyse the manufacturing process [108]. Possessing a vast amount of industrial data and domain knowledge, IFMs are able to realise knowledge retrieval and process analysis from historical industrial data and fault logs [109], thereby dynamically analysing and optimising the manufacturing process in industrial environments.

In conclusion, IFMs have been widely applied to various industrial domains to solve specific tasks. Ranging from IDMs adapted for industrial domains and ITMs tailored to specific industrial tasks, these models have been widely used in industrial domains such as electronics, steel, textiles, robotics, industrial control, construction, energy, automotive, and aerospace to address specific tasks. Some IDMs and ITMs for industrial domains are shown in Table 2.

Electronics: Electronics IFMs integrate consumer usage patterns, and supply chain information to improve product quality, resource management, and user experience [110]. Haier HomeGPT autonomously adapts to user behaviours and offers proactive services. OPPO AndesGPT leverages RAG technologies and agent-based logic to develop personal service. BOE XianShi enhances defect detection with high-fidelity imaging and computer vision foundation models. Nvidia designs ChipNeMo to assist engineers in completing tasks related to chip design. Electronics IFMs aim to address rapid product cycles, diverse consumer requirements, and complex global supply chains, enabling swift adaptation and better-informed decision-making.

Steel: Steel IFMs are developed to leverage operation logs, historical events, domain guidelines, and sensor data to stabilise production and ensure product quality [111,112]. Baowu Steel IFM optimises machining processes, such as rolling parameters, to accelerate production line adjustments, while ZENITH Steel IFM employs historical failure analysis for timely fault diagnosis. ANSTeel IFM uses computer vision technologies to identify real-time surface defect detection. These capabilities help manage parameter variations for different steels, maintain consistent output standards, and improve operational efficiency.

Textile: Textile IFMs combine market analysis, consumer preferences, and production factors to achieve agile product development, efficient inventory management, and responsive customer service [113,114]. Semir IFM uses knowledge bases to answer queries rapidly, enhancing customer support. ANTA IFM leverages AI-generated content to align product designs with evolving consumer tastes, and BOSIDENG IFM applies "AIoT + Large Model" approaches to fine-tune inventory levels. They can promote rapid iteration and provide stable supply chains, ensuring that textile manufacturers can swiftly respond to shifting market conditions.

Robotics: IFMs fuse visual inputs, sensor data, and language instructions to guide intelligent and context-sensitive robotic actions. VoxPoser translates textual policies into feasible movements via vision language models (VLMs) [115]. ManipLLM [102] uses RGB images and textual prompts for trajectory prediction. Besides, it integrates embodied sensor data with language models to refine robotic perception. Robotics IFMs emphasise zero-shot or few-shot adaptability, continuous learning, and robust motion planning, enabling robots to efficiently operate in dynamic industrial environments.

Industrial Control: Industrial control IFMs blend operational data, engineering principles, and equipment specifications to support predictive and prescriptive decision-making [116,117]. SUPCON-TPT employs simulation and time-series analysis to enhance flow manufacturing. Siemens Industrial Copilot integrates industrial expertise with code generation for efficient PLC programming. Yuanshan AI uses retrieval-augmented generation and agent-based logic to improve process control. Industrial Control IFMs can reduce downtime and detect faults before escalation to maintain stable production.

Construction: Construction IFMs handle engineering documents, regulatory codes, and spatial–temporal data to support safer and more efficient building projects [118,119]. Construction-GPT accelerates information retrieval, aiding on-site decision-making under tight schedules. Regulation and standards knowledge model are developed to ensure compliance with multiple building codes and safety protocols [51]. AecGPT and UrbanGPT [120] blend design criteria, geospatial insights, and historical data for blueprint validation and workflow optimisation. Construction IFMs can manage complex processes and enable data-driven project coordination across varied construction stages.

Energy: Energy IFMs analyse operational parameters, resource availability, and environmental conditions to stabilise output, forecast demand, and ensure safety [121,122]. Antelope Energy Model and Kunlun Model are developed to optimise energy production planning and forecasting accuracy. Big Watt processes extensive grid data for early fault detection, while the Kunlun Model refines meteorological analysis to improve renewable energy scheduling. Energy IFMs enhance predictive analytics, balanced resource allocation, and environmentally responsible operation.

Table 2
Current IDMs and ITMs in industrial domains.

Domain	Name	Details	Functions
Electronics	Haier HomeGPT	HomeGPT integrates LFM and multimodal technologies to enhance understanding, perception, decision-making, and adaptability in smart home scenarios. It enables full-scenario intelligent interactions and solutions. It can achieve autonomous learning when facing unfamiliar issues. Besides, it also supports proactive house services. For example, it can actively recommend meal plans based on user preferences and conditions, ensuring a balanced and nutritious diet.	Intelligent interactions
	OPPO AndesGPT	As the core AI engine of OPPO, the AndesGPT leverages technologies such as Agent and RAG to comprehensively empower OPPO smart devices. It continuously builds knowledge, memory, tools, and creative capabilities, delivering users a personalised and unique intelligent experience through AI agents.	Intelligent interactions
	BOE XianShi	The Xianshi can effectively address defect detection challenges that lack high-quality data and high learning cost problems. It helps improve the iteration and deployment efficiency of defect detection models by 10 times while boosting the efficiency of yield analysis and issue resolution by 5 times.	Quality management
	Nvidia ChipNeMo	ChipNeMo uses a large amount of electronic domain data for adaptive pre-training. The supervised fine-tuning with domain-specific tasks is used to improve its performance for chip design. Besides, a retrieval-augmented generation approach is also developed to improve the answer quality for chip design. ChipNeMo can assist engineers in completing tasks related to chip design.	Product design
Steel	Baowu Steel IFM	In the hot-rolling production line, engineers need to re-optimise over 300 process parameters whenever sizes or types of steels are adjusted. BAOWU STEEL IFM is developed to predict the optimal parameters that can reduce the time required to tune the hot-rolling production line. It can enhance the yield of finished steel plates and lower production costs.	Process optimisation
	ZENITH Steel IFM	ZENITH Steel IFM is designed for operations optimisation and maintenance in steel production. It integrates historical failure data and public industrial knowledge to provide solutions when faults occur. When steel production fails to meet targets, the system can quickly analyse data and identify causes of issues, making timely optimisation to the production schedule.	Predictive maintenance & Process optimisation
	ANSTeel IFM	ANSTeel IFM is utilised for quality inspection tasks in the steel industry, such as surface defect detection on steel and intelligent grading of scrap steel. High-resolution images collected from industrial cameras can quickly and accurately identify defects on steel surfaces and scrap steel, achieving an accuracy rate of over 95	Quality management
Textile	Semir IFM	Semir IFM has an AI-powered knowledge base, which includes a wide range of information across products, services, and processes. It makes the information easily accessible and shareable, helping the customer service team to answer customer questions more efficiently. Besides, it is also capable of independently answering simple questions.	Question answer
	ANTA IFM	Anta IFM utilises AI-generated content in the product development process to achieve more accurate insights into the requirements of users, design workflow optimisation, and product development. It can provide valuable suggestions for engineers about the style of clothes.	Product design
	BOSIDENG IFM	BOSIDENG has designed an "AIoT + Large Model" solution to analyse offline store inventory data and customer behaviour precisely. It optimises inventory management and product restocking strategies to boost store sales. By enhancing the intelligence of the decision-making process, it has significantly improved sales performance.	Data analysis
Robotics	VoxPoser	VoxPoser is a general robot control framework proposed by Fei-Fei Li at Stanford University. It utilises LLMs and VLMs to analyse natural language and make inferences, providing instructions and constraints based on the analysis. The composed value maps are then used in a model-based planning framework to zero-shot synthesise closed-loop robot trajectories with robustness to dynamic perturbations.	Intelligent control
	ManipLLM	ManipLLM is designed for Embodied AI. It uses RGB images and text prompts to predict trajectories of end-effectors through a chain of thought processes. Once initial contact is made, an active adaptation policy will be employed to plan the upcoming waypoints in a closed-loop manner.	Intelligent control
	PaLM-E	PaLM-E is an Embodied Model. It directly integrates real-world continuous sensor data into language models, establishing a connection between words and percepts. The inputs of PaLM-E consist of multimodal sentences that combine visual information, continuous state estimations, and textual inputs. It outputs the instructions and commands for robot control.	Intelligent control
Industrial control	SUPCON-Time-series Pre-trained Transformer	TPT combines simulation and prediction capabilities to support flow manufacturing control in different industrial scenarios. Massive industrial data from various production, operations, processes, and equipment are used to train it. It can achieve high-precision, high-reliability and closed-loop applications across different industrial scenarios, thereby significantly enhancing the efficiency of industrial applications.	Predictive maintenance
	Siemens Industrial Copilot	Siemens developed Industrial Copilot by leveraging extensive industrial knowledge, real-world cases, and domain-specific expertise. It has been trained to understand natural language for PLC code generation and HMI graphical interface creation by integrating the ladder diagram, instruction list, and hardware. It can reduce repetitive tasks, enhance production efficiency, and shorten development time.	Code generation

(*continued on next page*)

Table 2 (*continued*).

	Yuanshan AI	YuanShan Al adopts a hybrid artificial intelligence approach, integrating RAG, Agent, and multimodal technologies in intelligent industrial control scenarios. It aims to achieve effective advanced process control, manufacturing process optimisation and process improvement of industrial control.	Process optimisation
Construction	Construction-GPT	Construction-GPT integrates IFM with industry-specific standards, drawings, and technical documents. Construction-GPT can support a quick dialogue-based query process, which reduces the time for technical references, enabling engineers and project managers to make faster, data-driven decisions directly at worksites.	Content generation
	AecGPT	AecGPT is developed based on Glodon's industry-focused AI tools by integrating IFMs with a broad range of construction-specific data—standards, drawings, and regulations. It ensures comprehensive domain knowledge and seamless adaptation to various construction workflows. The key functions are automating repetitive tasks, enabling rapid retrieval of technical information, and generating professional-level content.	Decision making
	UrbanGPT	UrbanGPT is developed by integrating a spatio-temporal dependency encoder into LLMs. The instruction-tuning paradigm is employed to enhance generalisation across diverse spatio-temporal learning scenarios. It aims to accurately predict and analyse various urban dynamics over time and space, support zero-shot spatio-temporal prediction tasks, and provide insightful, data-driven guidance for urban planning, resource allocation, and policy-making.	Decision making
	CivilGPT	CivilGPT is a domain-specific IFM developed by integrating large-scale professional knowledge bases such as multi-course curricula, textbooks, standards, and exam questions into a civil engineering knowledge graph. It can provide personalised learning paths, intelligent Q&A, and context-aware decision support in civil engineering education and practice.	Question & Answer
Energy	Kunlun Model	Kunlun Model is designed to integrate industrial data with scenarios on a unified AI platform. It enhances domain-specific intelligence across exploration, refining, sales, and equipment manufacturing. Additionally, it enables multimodal human-machine interaction and supports continuous model optimisation through data feedback.	Process optimisation
	Big Watt	South China Power integrates large-scale domestic AI to construct independent and controllable electric power models. The key functions are to enhance automation and accuracy in power grid operations, such as detecting line defects, improving inspection efficiency by identifying multiple types of equipment flaws, and providing real-time assistance in diverse application scenarios.	Predictive maintenance
	Antelope Energy Model	Antelope Energy is developed by leveraging the iFlytek Spark V4.0 large model as the core technology to integrate extensive energy industrial datasets and domain knowledge. It is tailored to fit diverse energy scenarios like wind, solar, hydro, thermal, and nuclear. It can handle multimodal information, generate specialised content, conduct knowledge-based Q&A, and forecast renewable power.	Process optimisation
Automotive	BMW Assistant	The BMW Assistant is developed by integrating Amazon's Alexa IFM into BMW's system and AI platform. It features augmented reality visualisation through XREAL Air 2 AR glasses and enhanced connectivity. Its key functions include personalised voice assistance for vehicle-related queries, real-time AR navigation and hazard warnings, and seamless integration of entertainment features.	Decision making
	Xpeng Tianji	Xpeng Tianji is built on LLM with extensive real-world driving data. It integrates a map-free approach with AI Eagle Eye vision to achieve L3-level autonomous driving capabilities. Its key functions include delivering advanced autonomous driving across all vehicle lines without additional cost, enabling adaptive chassis control, and personalised cockpit features.	Decision making
	BYD Xuanji	BYD Xuanji leverages the integration of electrification and multimodal technologies to develop IFM. Its core capabilities include end-to-end intelligent vehicle applications, spanning perception, decision-making, and precise control. It supports advanced autonomous driving and offers automotive parking. Additionally, Xuanji delivers personalised in-cabin experiences and incorporates adaptive systems to enhance operational efficiency comprehensively.	Decision making
Aerospace	Aerospace-Baidu Wenxin Model	It is developed by integrating Wenxin IFM with specialised aerospace-sector datasets. Employing advanced computational techniques, domain-specific knowledge to enhance the comprehension and processing of complex deep-space information. The key functions include automated data analysis, knowledge integration, and context-specific generation for deep-space exploration.	Question answer
	NASA Embedded Knowledge + ChatGPT	The model is developed by integrating GPT-4 with a Neo4j-based knowledge graph for NASA's roadmap, combining retrieval-augmented generation and graph-driven queries. The key functions are automated requirement analysis, relation prediction, and enhanced data organisation, enabling more precise decision-making.	Decision making

Automotive: Automotive IFMs integrate sensor data, engineering specifications, and operational logs to optimise overall driving experiences [123]. BMW Assistant leverages IFMs frameworks for responsive in-vehicle support, guiding drivers through dynamic traffic conditions while optimising vehicle performance [124]. Xpeng Tianji applies vision-based reasoning to refine autonomous driving. BYD Xuanji incorporates multimodal inputs, including camera images and radar signals, to achieve reliable, efficient, autonomous navigation. Automotive IFMs facilitate informed decision-making and maintain compliance with rapidly evolving vehicular standards.

Aerospace: Aerospace IFMs employ domain-specific datasets, engineering schemas, and retrieval-augmented reasoning to manage mission planning, data organisation, and regulatory adherence. Aerospace-Baidu Wenxin Large Model integrates technical documentation, flight data, and environmental parameters to guide operational decisions in advanced aerospace applications. NASA Embedded Knowledge Graph+ ChatGPT utilises structured queries and domain-tailored retrieval techniques to streamline resource allocation, compliance checks, and knowledge management [125]. Aerospace IFMs enhance safety protocols, optimise route planning, and provide decision support that meets the

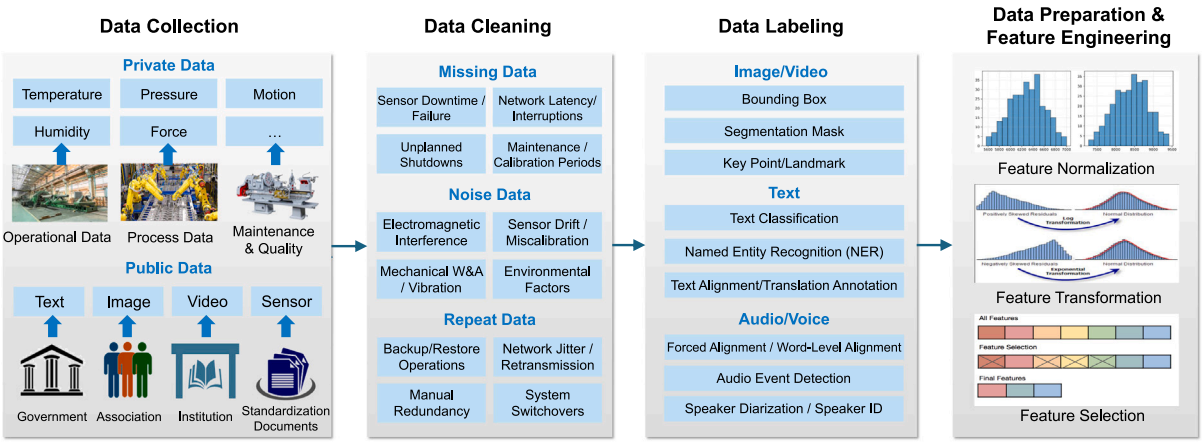

Fig. 5. Data analytics approach for IFMs.

stringent demands of aerospace exploration, ensuring that missions are aligned with intricate engineering constraints [90].

4. Key technologies of IFMs

4.1. Data analytics of IFMs

Data analytics integrates large-scale and heterogeneous data to extract meaningful features, which is the foundation of IFMs. It includes data acquisition, data cleaning, data labelling, data preparation and feature engineering, as illustrated in Fig. 5.

4.1.1. Data acquisition

Data acquisition constitutes the foundation for advanced analytics and predictive modelling of IFMs [126]. Industrial data encompasses enterprise-private and public data as their different sources, accessibility, sensitivity, and usage.

Public data refers to industrial data accessible to everyone and can be freely shared [127]. It does not include critical business information or private details. Common types of public data include four primary groups: (1) Government and regulatory data: regulations, production data, standards, growth trends and other statistics made available through government agencies. (2) Association data: published market trends, technical standards, and best-practice guidelines, which are often presented as textual documents or reports [128]. (3) Standardisation documents: text-heavy references such as invention patents and technical reports that are used to improve implementation and drive technological innovation within industries. (4) Institutional and academic research outputs: peer-reviewed articles, open-access datasets, and benchmarking studies from universities or research institutes. These datasets cover different tasks (e.g., defect detection, fault diagnosis) across various industrial scenarios, presented in multiple formats, including text, images, videos, and audio [129].

Private data includes an industrial dataset, which contains sensitive information restricted to authorised users. It often involves technology details or customer-related information, which are crucial for maintaining a competitive edge [130]. Private data typically contains three major categories: (1) Operational data: real-time sensor readings (e.g., vibration amplitude, temperature, force) and industrial IoT capture continuous numeric signals related to machine health [131]. (2) Process data: production schedules, assembly-line metrics, outputs from supervisory control, and data acquisition platforms typically encompass numerical records. (3) Maintenance records: equipment maintenance logs, fault reports, and quality inspection results stored as text documents (e.g., work-order summaries, spare-part usage reports) [132].

Private data often contains personally identifiable information, making privacy-preserving approaches crucial in industrial contexts. To ensure data privacy, secure multi-party computation (SMPC) can be employed to encrypt data at source nodes for collaborative analysis, ensuring that original data never leaves local environments. For instance, in supply chain analytics, manufacturers and suppliers can encrypt their respective data and use SMPC to jointly train predictive models, with only aggregated model updates shared. This approach prevents the exposure of private data while enabling collaborative model training. Differential privacy introduces controlled statistical noise into data to ensure that the presence of a single data record does not significantly affect model training, thereby mathematically guaranteeing the indistinguishability of individual privacy. For example, Laplace noise can be added to data before it is input into IFMs. The noise intensity is regulated by a privacy parameter ϵ. Lower ϵ values mean stronger privacy, but the data becomes less precise. Higher ϵ values allow the model to be more accurate but offer less privacy protection. It is a trade-off between keeping data private and making sure the model works well.

4.1.2. Data cleaning

Industrial raw data is frequently compromised by a variety of issues, including missing values, noisy data, and redundant data, each of which can critically undermine the reliability and robustness of downstream analyses [133,134].

Missing data may arise from sensor downtime, network latency, or equipment shutdowns [135]. Advanced imputation techniques, such as basic statistical interpolation (e.g., linear or spline interpolation) and more elaborate approaches like multiple imputation, can help maintain temporal coherence and signal integrity. Besides, domain knowledge about normal ranges or standards can be utilised to guide how and when to interpolate missing values [136].

Noise data is a common issue in industrial data analytics, as disturbances, sensor errors, or mechanical wear can introduce spurious spikes or drifts [137]. Filtering algorithms (e.g., moving average, Butterworth filters) can smooth out high-frequency fluctuations, while wavelet-based de-noising methods [138] can systematically identify and remove transient anomalies while preserving critical event markers. For specialised applications such as vibration analysis in rotating machinery, frequency-domain techniques (e.g., Fourier transform) filter out fault signals from noise [139].

Redundant or duplicated data may appear when data acquisition systems overlap [140]. Though some levels of redundancy can serve as a backup measure, excess duplication risks will inflate data volumes and slow model training. Data integrity involves cross-verifying timestamps, sensor IDs, and recorded parameters to pinpoint duplicated rows or conflicting data segments. By implementing version control

systems and robust data-cleaning pipelines, industrial organisations ensure that their datasets remain coherent to reduce the likelihood of fake correlations.

4.1.3. Data labelling

Accurate labelling is critical for supervised model training, essential in industrial contexts, such as predictive maintenance, defect detection, and process optimisation [141].

In predictive maintenance, historical sensor data must explicitly indicate instances of abnormal wear, partial malfunctions, or breakdowns. These labelled instances from the positive examples can be used to train classification or regression models for prediction. When dealing with defect detection tasks, vision systems capture images or videos of products, which are then annotated with bounding boxes or segmentation masks that locate and describe flaws such as scratches and stains. However, high-quality labels can be time-consuming and expensive, particularly in large-scale industrial settings [142]. Contemporary data annotation pipelines increasingly rely on systematic labelling protocols and rigorous quality assurance to support large-scale machine learning applications [143]. In audio processing, audio event detection and speaker identification further highlight the need for meticulously defined labelling guidelines. Although crowd-based platforms can expedite these processes for high-volume, lower-complexity tasks, final validation by domain experts remains indispensable in specialised settings [144]. To ensure reliability across annotators, inter-annotator agreement metrics (e.g., Cohen's kappa, Fleiss's kappa) can be employed to quantify consensus levels and reveal potential sources of variance [145]. By integrating structured workflows, continuous verification, and domain-informed oversight, modern annotation pipelines can effectively align labelling outputs with the stringent quality demands of advanced machine learning systems [146]. Where data privacy is a concern, labelling processes may also be conducted using anonymised datasets or secure platforms that limit the exposure of sensitive details to annotators. Techniques such as differential privacy can ensure that statistical aggregates derived from labelled data do not reveal individual-level confidential information.

4.1.4. Feature engineering

Following the data cleaning and labelling stages, datasets are refined for model training through targeted preparation and feature extraction [147].

Feature normalisation can mitigate scale discrepancies among sensor readings [148]. For instance, the temperature may vary from 0 °C to 100 °C, vibration amplitude is measured in microns, and pressure spans several orders of magnitude in Pascals. Techniques like min–max scaling and Z-score normalisation convert data to a uniform numerical range, thereby improving model training efficiency and accuracy [149]. Feature transformation enhances the model's capacity to identify underlying patterns. In the context of rotating machinery diagnostics, Fourier transforms can isolate characteristic frequency bands indicative of mechanical faults such as unbalance and misalignment, while wavelet transforms are able to effectively localise transient spikes or abrupt shifts in time-series data. Feature selection mitigates dimensionality issues, particularly when numerous parameters (e.g., temperature, flow rate, torque, and acoustic signals) are recorded [150]. Iterative wrapper and embedded methods evaluate feature importance and systematically eliminate less informative variables, thereby reducing computational demands while preserving predictive accuracy. Data privacy can be integrated into feature engineering by excluding or anonymising sensitive data, while secure computation and encryption allow multiple stakeholders to contribute without disclosing proprietary information. These measures enable robust IFMs that fully leverage large-scale data while maintaining strict privacy standards.

4.2. Development technologies for IFMs

This section presents the key development technologies for IFMs. It includes an introduction to pre-trained foundation models, instructions on training IFMs using foundation models as the backbone, and instructions on adopting IFMs for specific industrial applications.

4.2.1. Foundation models

The idea of the "foundation model" was initially proposed by Bommasani et al. [151] at Stanford's Human-Centred AI initiative. These models are broadly characterised as large-scale frameworks developed through self-supervised or semi-supervised learning, making them adaptable for numerous downstream tasks. This shift departs from narrowly focused models, favouring general-purpose systems trained once and that can handle diverse applications. The approach facilitates faster model customisation, enhances performance across different contexts, and demonstrates unique emergent capabilities stemming from training on vast datasets [45,152].

The term "industrial foundation model" lacks a universally standardised definition but generally refers to extensive AI models, often deep learning-based, designed for large-scale, practical applications in industries like manufacturing and robotics. These models are built using vast datasets and are tailored to address complex, real-world challenges. Frequently, IFMs rely on pre-trained foundation models as their backbone, which are either fine-tuned or further trained to suit specific industrial needs. Pre-trained foundation models have demonstrated exceptional capabilities in feature representation learning across tasks such as text analysis [153,154], image processing [155], and graph-based classification [156]. These models leverage large-scale datasets for initial training and efficiently adapt to smaller, related tasks, enabling swift data analysis. With a focus on industrial and manufacturing applications, this study examines pre-trained models in domains such as vision, language, and their intersection, alongside exploring their impact in areas like NLP, CV [157], and graph learning (GL) [158].

In 2021, the term "foundation model" was introduced to describe models trained on vast datasets that are capable of being adapted to numerous downstream tasks [151]. Three key features distinguish these models:

- In-context learning: This allows models to perform new tasks using only a few examples, eliminating the need for additional training or fine-tuning.
- Scaling laws: These indicate that performance continues to improve as the size of the dataset, computational power, and model complexity increase.
- Homogenisation: This feature enables specific architectures to process different data types and tasks within a unified framework.

These characteristics make foundation models highly versatile for a variety of applications.

This section discusses pre-trained foundation models categorised by input type: the Transformer, as well as pre-trained models for language, vision, vision-language, and multimodality.

4.2.2. Pre-trained foundation models

Transformer: The Transformer architecture serves as the foundation for numerous advanced models, including GPT-3 [159], DALL-E-2 [160], Codex [161], and Gopher [162]. It is developed to address the challenges of traditional models like RNNs in managing variable-length sequences and maintaining context. At its core, the Transformer relies on a self-attention mechanism, enabling the model to effectively focus on different parts of an input sequence.

The architecture includes two primary components: an encoder and a decoder. The encoder processes the input sequence to generate hidden representations, while the decoder uses these representations to produce the output sequence. Each layer in the encoder and decoder includes multi-head attention mechanisms and feed-forward neural

networks. The multi-head attention mechanism is pivotal, as it assigns varying levels of importance to tokens based on their relevance. This allows the model to capture long-term dependencies better and improve performance across various NLP tasks.

The Transformer's design is also highly parallelisable, enabling efficient computation and reducing reliance on inductive biases [76]. This property makes it suitable for large-scale pre-training, allowing Transformer-based models to adapt seamlessly to diverse downstream tasks.

Pre-trained language foundation model: LLMs refer to large-scale deep neural networks trained on massive datasets and are the foundation for modern NLP. A significant advancement in this field is the introduction of Transformer-based architectures, which dominate pre-trained language models' backbone due to their efficiency and scalability. Transformers, particularly in recent LLMs, often use autoregressive decoder-only designs consisting of multiple Transformer blocks trained to predict the next token. These autoregressive models, such as GPT-3 [163] and OPT [164], follow a left-to-right language modelling approach, making them highly effective for generative tasks. In contrast, masked language models, like BERT [165] and RoBERTa [166], focus on predicting masked tokens within a sentence and excel at extracting contextual representations for downstream tasks.

The datasets powering these models are of Internet-scale. For example, GPT-3's [163] use of 45TB of filtered data from CommonCrawl, spanning from 2016 to 2019, and Codex's [167] training on a snapshot of 54 million public software repositories from GitHub to enable programming capabilities. These pre-trained models generalise across diverse domains and handle multiple data formats, such as natural language, programming languages, and structured data like JSON and YAML.

LLMs are versatile in their input–output capabilities, primarily supporting two transformations: Language → Language and Language → Latent. For Language → Language, models like GPT-3 [163], LLaMA [76,76], and others facilitate complex reasoning tasks through techniques such as Chain of Thought (CoT) prompting [168,169]. These models are not restricted to natural language, they can also process programming code and other structured formats. In contrast, Language → Latent transformations are exemplified by models like BERT [165] and RoBERTa [166], which map language into latent space representations. These embeddings can be used for sentence similarity, information retrieval, and other tasks by leveraging intermediate-layer features.

Language models offer additional advantages in robotics. They provide robots with human-like common sense, enabling them to perform motion planning and object recognition tasks. Through in-context learning, robots can adapt to new tasks with minimal examples. However, challenges like hallucination – where models generate incorrect outputs – highlight the need for robust mechanisms to ensure reliability. Overall, Transformer-based LLMs, whether autoregressive or masked, continue to revolutionise natural language processing and beyond, driving advancements in generative, retrieval, and multimodal applications.

Pre-trained vision foundation models: Large Vision Models (LVMs), which serve as foundation models for images, can be broadly classified based on their input and output relationships into two categories:

• Vision → Latent: These models are designed to map visual data from diverse sources into a latent space, effectively compressing image information into lower-dimensional vectors that capture high-level features. Examples include R3M [170] and VC-1 [171], which focus on extracting meaningful representations from images. This type of learning is often self-supervised, enabling the construction of large datasets with minimal manual effort.

• Vision → Recognition: This category involves performing tasks such as semantic segmentation, instance segmentation, and object detection on images. Recognition tasks typically rely on language labels to classify or annotate images. Models like Segment Anything [33] excel in image segmentation, allowing for precise segmentation of entire images or specified regions using points or bounding boxes. Building on this, more advanced models like Tracking Anything [172] and Faster Segment Anything [173] have been introduced for enhanced applications.

While some models combine aspects of Vision → Vision or Latent → Vision, effectively utilising them for diverse tasks often requires conditioning on language. Vision → Recognition tasks, in particular, are more labour-intensive due to the need for manually labelled datasets, unlike the more scalable self-supervised approaches of Vision → Latent tasks. The following section will explore the integration of vision and language in these models, showcasing their broader applications.

Pre-trained multimodal foundation models: Recent advancements in Multimodal Large Language Models (MLLMs) have been explored extensively by Zhang et al. [174], who examined neural networks that extend LLMs with multimodal capabilities. Their study evaluates the performance of significant MLLMs across 18 vision-language benchmarks and provides an overview of MLLMs architectures. As an example, IFMs built upon metaverses are introduced to support natural interactions and intelligent operations between humans and machines. They function as the operating systems of industrial parallel machines, offering persistent data resources and diverse scenarios for control and management tasks. In this context, IFMs integrate vision foundation models, language foundation models, and operational foundation models. This unique combination is designed to handle resource management within industrial parallel machines and deliver comprehensive support for industrial workflows [175]. However, their analysis does not address the Type-D 3.4 multimodal architecture, a novel and increasingly popular approach for creating flexible any-to-any modality models. Similarly, Yin et al. [176] offer an in-depth exploration of typical multimodal architectures, including pre-training, fine-tuning, alignment methods, and data usage. Yet, their work also fails to address Type-D architecture. Caffagni et al. [177] present a detailed review of multimodal architectures, summarising key components such as LLM variants, vision encoders, and adapters for vision-to-language connections. Despite their broad discussion of domain-specific applications like document analysis, medical imaging, autonomous driving, and embodied AI, Type-D multimodal architecture remains unmentioned.

Efficient MLLMs follow a standard framework comprising three core modules: a visual encoder GGG to process visual inputs, a pre-trained language model for multimodal reasoning, and a visual-language projector PPP to align the two modalities. MLLMs' optimisation efforts focus on processing high-resolution images, compressing vision tokens, and employing lightweight language models to improve efficiency. Key strategies include compact architectures, efficient structures, and token compression mechanisms.

Built on the above foundation models, research efforts of IFMs on real-world applications have been widely explored in recent years. The key applications in manufacturing and robotics are summarised below, and they are powered by large-scale pre-trained models tailored for industrial data modalities such as vision, language, CAD, IoT, and robot trajectories:

- *Visual inspection & defect detection:* IFMs trained on multimodal industrial data (e.g., vision-language pairs, CAD renderings) can generalise to novel defects and parts without task-specific retraining [178].
- *General-Purpose skill learning*: Robotics Foundation Models learn skills across diverse tasks (e.g., pick-and-place, assembly) using language-conditioned policies or video-language datasets [179, 180].
- *Predictive maintenance & industrial time series modelling:* Foundation models for time-series and sensor data can detect early signs of equipment failure across different machine types [181].

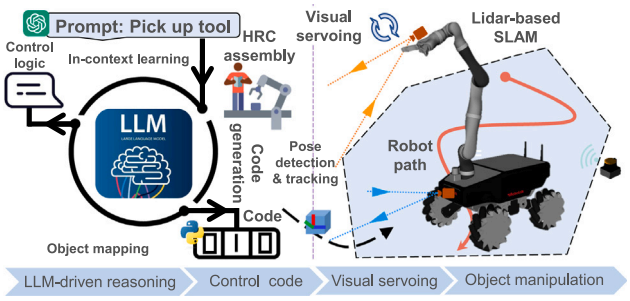

Fig. 6. LLM-enabled HRC assembly [192].

- *CAD & CAM understanding:* CAD foundation models understand and generate 3D shapes, assist in design automation, and interpret design intents via natural language or sketch inputs [182,183].
- *Digital twin modelling and simulation:* Foundation models power digital twins for real-time process modelling, optimisation, and simulation from historical plant data and structured documents, such as simulation of energy systems, production lines using multimodal industrial data [184–186].
- *Assembly planning & task understanding:* IFMs assist in understanding part relations, assembly sequences, and affordances from multi-modal input (text, image, 3D scans), and the typical examples include zero-shot task understanding and automated plan generation for new products [187–189].
- *Natural language interfaces for machines or robots:* Multimodal foundation models allow users to interact with machines using language, e.g., "assemble this part", or "inspect for scratches [190]. For example, Fig. 6 shows an overview of LLM-enabled HRC assembly. It employed a large language model-driven reasoning method for text-based assembly task description and robot control commands interpretation and execution [191].

4.2.3. Learning methods for pre-trained foundation models

Pre-trained foundation models represent a transformative shift in machine learning, offering a scalable and versatile framework for diverse tasks across domains. These models leverage a variety of learning paradigms, each contributing uniquely to their capability to generalise and adapt. The primary learning methods include supervised, semi-supervised, weakly supervised, self-supervised, and reinforcement learning, often applied in pre-training, fine-tuning, and transfer learning stages [193]. Table 3 lists the comparison of these learning methods.

Supervised learning: Supervised learning [194] forms the foundation of many traditional machine-learning approaches by utilising datasets with explicit input–output pairs. This method is highly effective for tasks requiring precision, such as image classification, object detection, and text translation. However, its reliance on large, high-quality labelled datasets can pose significant challenges in cost and scalability.

Semi-Supervised learning: To address the limitations of supervised learning, semi-supervised learning [195] combines a small amount of labelled data with a larger pool of unlabelled data. Techniques like pseudo-labelling and consistency regularisation enable models to utilise unlabelled data effectively. This approach is beneficial when obtaining labelled data is expensive or impractical, such as in medical imaging or resource-constrained natural language applications.

Weakly supervised learning: Weakly supervised learning [196] reduces the dependence on perfectly labelled datasets by using noisy, incomplete, or imprecise annotations. Common in medical diagnostics and remote sensing applications, this method relies on strategies such as label smoothing and noise-robust training to handle label uncertainty while extracting meaningful insights.

Self-supervised learning: Self-supervised learning [197] allows models to learn directly from the data structure by defining surrogate tasks, eliminating the need for manual labelling. Methods like masked language modelling in BERT and contrastive learning in vision models (e.g., SimCLR) exemplify this paradigm. This approach is a cornerstone of pre-training foundation models, enabling them to acquire robust, generalisable representations suitable for downstream tasks.

Federal learning: Federal learning [198] is a distributed machine learning approach that enables multiple parties to collaboratively train a shared global model while keeping their data localised. It offers a promising solution for privacy-preserving, distributed model training by allowing collaborative learning without centralising raw data. In industrial tasks, a federated learning approach can be used to protect the data privacy of industrial big data. The model can be trained locally using private data first, and then transferred to the cloud for collaborative training to build a complete model.

Reinforcement learning: Reinforcement learning [199] trains models through interaction with an environment, using rewards and penalties to optimise decision-making. This paradigm is especially valuable for sequential tasks like robotics, game-playing, and autonomous systems. Combined with pre-trained foundation models, reinforcement learning enhances their reasoning and adaptability in dynamic, complex environments.

Explainable AI (XAI): Deep learning models are often regarded as "black boxes" due to their complex network structures and a large number of parameters. This restricts their applications in industrial scenarios with high requirements for safety and decision-making transparency. XAI aims to mitigate this black-box nature and understand the reasoning behind the decisions of AI models. Current research on XAI focuses on data explainability, model explainability, and post-hoc explainability. Data explainability aims to gain a better understanding of the datasets before training the model. Common methods include exploratory data analysis, explainable feature engineering, dataset description standardisation, and knowledge graphs. Model explainability aims to create deep learning models that are naturally more understandable. Hybrid explainable models, such as Deep Weighted Averaging Classifiers, Contextual Explanation Networks, and Neural-Symbolic models, can visualise the weights of convolutional kernels or the propagation process. This helps to present the internal information of the model and reveal its decision-making mechanism. Besides, the regularisation techniques or decision trees can also be utilised to enhance the explainability of models. Post-hoc explainability elucidates significant features to analyse the decision-making process of deep learning models. The Gradient-weighted Class Activation Mapping (Grad-CAM) can generate high-resolution activation maps and mark the most critical parts by introducing gradient information. There are also methods that aim to quantify the contribution of each feature to the model prediction, thereby improving the explainability of deep learning models.

Stages of learning for foundation models: Foundation models typically undergo three stages of learning: pre-training, fine-tuning, and transfer learning [200].

• Pre-training involves training on large-scale datasets, often using self-supervised or semi-supervised techniques, to learn general-purpose representations.

• Fine-tuning tailors these pre-trained representations to specific downstream tasks using smaller, task-specific datasets.

• Transfer learning leverages the representations learned during pre-training to adapt models efficiently to new tasks or domains. It often requires minimal labelled data.

These learning methods and stages empower pre-trained foundation models to excel across diverse applications. These models balance efficiency, scalability, and adaptability by integrating multiple paradigms, making them indispensable in modern AI systems.

Table 3
Comparison of the selected learning methods.

Aspect	Supervised learning	Semi-supervised learning	Weakly-supervised learning	Self-supervised learning	Reinforcement learning
Definition	Learning from fully labelled data.	Learning from a combination of labelled and a larger amount of unlabelled data.	Learning from imprecise, incomplete, or noisy labels.	Learning by generating labels from input data without explicit human annotations.	Learning through trial-and-error by interacting with an environment to maximise cumulative rewards.
Type of data	Fully labelled dataset.	Small labelled dataset and large unlabelled dataset.	Data with weak or noisy labels (e.g., partial annotations, class labels instead of instance labels).	Unlabelled data with the inherent structure for defining tasks.	Interaction data with states, actions, and rewards from the environment.
Goal	Predict correct labels for new data.	Use unlabelled data to improve prediction accuracy.	Learn robust models from weakly labelled data.	Learn rich data representations or solve pretext tasks.	Learn an optimal policy for decision-making or control.
Dependency on labels	High—requires labels for all training examples.	Moderate—requires fewer labels compared to supervised learning.	Moderate—relies on weak labels.	Low—no manual labels required.	Depends on reward signals, not explicit labels.
Advantages	High accuracy and reliable predictions when data is fully labelled.	Reduces labelling costs by utilising unlabelled data.	Handles scenarios where exact labelling is impractical or expensive.	Learns features that generalise well across tasks.	Solves sequential decision-making problems; handles dynamic environments.
Challenges	Expensive and time-consuming to label large datasets.	Requires careful balance between labelled and unlabelled data.	Performance depends on the quality of weak labels.	Designing effective pretext tasks and architectures.	Exploration vs. exploitation trade-off; sparse or delayed rewards can be challenging.
Key examples	Image classification, object detection with labelled datasets.	Semi-supervised image segmentation, text classification.	Learning from tags, bounding boxes instead of pixel-wise annotations.	Contrastive learning, autoencoders, or masked token prediction (e.g., BERT).	Robot control, game playing (e.g., AlphaGo), autonomous driving.
Commonality	Data-driven approaches use optimisation techniques to minimise errors or maximise objectives.	Exploits data to learn a mapping from input to output or an effective policy based on task requirements.	Learning Paradigm: Uses available data with task-specific objectives to train models in a defined domain.		

4.2.4. Industrial domain models

The IFMs represent a transformative AI technology tailored for industrial contexts. It utilises vast, industry-specific datasets and template libraries encompassing sensor data, events, asset details, and operational insights to deliver precise and actionable intelligence. By streamlining complex industrial processes, IFMs provide valuable predictions and insights, making them essential for enhancing manufacturing, maintenance, and process optimisation.

To further elevate data intelligence across industries, researchers at Microsoft Research Asia have introduced the concept of IFMs. Their strategy involves adapting IFMs through post-training on industry-specific data science tasks, embedding domain-specific expertise, and refining in-context learning capabilities. This approach aims to develop IFMs that excel in diverse tasks, enabling predictive and logical reasoning tailored to industrial needs while extracting transferable knowledge across various domains. SymphonyAI has announced one of the first industrial LLMs to accelerate large-scale industrial transformation, underscoring the growing role of such models in reshaping industry operations.

The Industrial LLM is built using one of the most extensive industrial datasets globally, encompassing 1.2 billion tokens, 3 trillion data points, over 500,000 machine tests, 150,000 components, and 80,000 unique assets. This vast training foundation enables it to harness predictive and generative AI to enhance operational efficiency, productivity, and profitability by providing operators with contextualised insights for faster, better-informed decisions. The model delivers actionable, context-aware data up to 90% faster than conventional systems, significantly improving response times.

Hosted on Microsoft Azure, the Industrial LLM seamlessly integrates and contextualises manufacturing operation data at all scales, from individual equipment to multi-plant global operations. It serves as a stand-alone intelligence system for addressing asset performance and reliability questions or can be connected with downstream systems and plant data sources. Both deployment modes facilitate real-time, measurable business outcomes while cultivating a more knowledgeable and interconnected workforce.

By leveraging diverse data sources, such as events, sensor readings, asset details, work orders, warranty information, product documentation, troubleshooting manuals, and maintenance reports, the Industrial LLM empowers manufacturers to derive meaningful insights. Its self-learning capabilities allow it to adapt dynamically to new data and operational changes, ensuring it stays relevant in fast-paced environments. Additionally, the Industrial LLM makes years of domain expertise accessible to operators and plant managers, enabling even less experienced users to effectively address complex challenges and drive improved outcomes across manufacturing and maintenance workflows.

The Industrial LLM is among the first large language models explicitly tailored for industrial applications. It is trained on extensive proprietary datasets and a carefully curated knowledge base to handle tasks critical to industrial users. These tasks include diagnosing machine conditions, providing prescriptive recommendations, and addressing inquiries about specific fault scenarios, test procedures, maintenance workflows, manufacturing processes, and industrial standards.

Currently offered in a private preview, the Industrial LLM enables developers to create custom industrial applications through its API. It will be accessible via the Microsoft Teams AI Library and as a model

Table 4
Key differences between LLMs and IFMs [175].

Aspect	LLMs	IFMs
Purpose	Focused on natural language understanding and generation.	Tailored for specific industry applications (e.g., healthcare, finance, manufacturing).
Data	Trained on general-purpose datasets, often web-scraped text.	Trained on domain-specific data with targeted datasets relevant to the industry.
Capabilities	Broad linguistic and reasoning abilities across many domains.	Deep expertise in narrow, specialised areas of industry.
Training scope	Generalised to handle diverse tasks like summarisation, translation, Q&A.	Focused on solving specific industry challenges, e.g., detecting anomalies in machinery.
Model size	Typically very large (e.g., GPT-4, LLaMA 2, exceeding 100B parameters).	Smaller to moderate size, often optimised for specific tasks or environments.
Customisability	Requires fine-tuning or prompting for domain-specific applications.	Often pre-customised for industry needs, requiring less additional tuning.
Deployment	Use in various general-purpose applications, including chatbots and content creation.	Integrated into industrial workflows, systems, or IoT environments.
Real-time usage	Prioritises human interaction and response generation.	Often designed for automation, predictive maintenance, or real-time decision-making.
Regulatory concerns	Handles generic content with minimal regulatory constraints.	Must adhere to strict industry standards and regulations (e.g., FDA for healthcare, ISO for manufacturing).
Accessibility	Open access or via APIs for a broad audience.	Often restricted access, with deployment requiring industry-specific expertise.
Optimisation goals	Maximising linguistic understanding and creativity.	Optimised for efficiency, accuracy, and reliability in industrial tasks.

within the Model Catalog in Azure Machine Learning Studio. Additionally, it is positioned as a valuable educational resource for universities and colleges, helping to equip the next generation of professionals with skills in intelligent manufacturing.

Unlike general-purpose models like LLMs, ILMs are developed with a targeted focus, leveraging domain-specific data, optimisation strategies, and architectures. Table 4 summarises the difference between LLMs and IFMs, and the Key Characteristics of ILMs are summarised below:

Domain-Specific Training: ILMs are trained on datasets curated from specific industries. As one example, manufacturing models might rely on sensor data, CAD designs, or assembly line logs. This ensures the models are highly specialised and effective within their application area.

Targeted Optimisation: Models are optimised for tasks like anomaly detection, predictive maintenance, fraud detection, or risk analysis. Unlike LLMs, which focus on broad language capabilities, ILMs prioritise metrics such as accuracy, precision, and reliability in specific contexts.

Compliance with Regulations: ILMs are designed with these compliance standards, ensuring their predictions and decisions meet legal and ethical standards.

Integration with Industrial Workflows: ILMs are often embedded within existing systems, such as Enterprise Resource Planning (ERP) systems, IoT platforms, or SCADA (Supervisory Control and Data Acquisition) systems. Their outputs are actionable and align with real-time industrial processes.

Efficiency Over Model Size: ILMs prioritise computational efficiency and real-time inference over sheer size. They are often compact enough to be deployed on edge devices or in environments with limited computational resources.

4.3. Application technologies for IFMs

Application technologies are significant for integrating IFMs with manufacturing systems to address industrial tasks. They mainly include prompt engineering, RAG, and agent engineering. The application framework of IFMs for intelligent manufacturing is shown in Fig. 7.

4.3.1. Prompt engineering

Prompt engineering is a method for tailoring a IFM to perform specific tasks by incorporating task-specific cues into the model's input [201]. This approach has risen to prominence alongside large pre-trained models, collectively driving a significant shift in machine learning paradigms [202]. Recently, leveraging pre-trained models through prompts to adapt them to targeted tasks has become increasingly popular. Prompt engineering involves embedding hints with inputs to enable a pre-trained model to tackle new tasks using existing capabilities. However, prompt engineering goes beyond merely creating prompts, it involves diverse skills and techniques for utilising IFMs. It is crucial for understanding, leveraging, and expanding the potential of IFMs. This technique can enhance the safety of IFMs and extend their utility with external tools [203]. A prompt typically consists of several components: instructions specifying the task, context providing additional guidance, input data outlining the query, and an output indicator defining the desired response type or format.

Table 5 summarises the commonly used prompting techniques and their advantages and differences. These prompt techniques include zero-shot, few-shot, CoT prompting, self-consistency, generated knowledge prompting, prompt chaining, tree of thoughts, RAG, directional stimulus prompting, program-aided language models, ReAct, Reflexion, multimodal CoT, and graph prompting. Given the features of different prompting methods, some emphasise prompting with few or zero examples. In contrast, others consider including reasoning steps in the prompt, which mimic human thinking when solving problems, especially CoT. In addition, the background and external information can also facilitate the model's capabilities to get the final results. Given this, the prompt design proposes RAG and generated knowledge prompting techniques. Compared with single or simple input information for the prompting, multimodal input commands (e.g., image, text) are used to prompt the model for efficient reasoning and task planning. Table 5 gives a detailed introduction to these prompting techniques in terms of specifications, advantages, and differences, which helps choose which prompt technique for ILMs. This can be a good guide to adopt a method for specific cases.

Table 6 summarised representative examples of the popular LLMs using the prompt methods, where these LLMs include ChatGPT, GPT-4, Claude 3, LLaMA, Flan, Gemini, Code Llama, Mistral, Phi-2, Sora, OLMo, Grok-1, and Mixtral. Here, some commonly used LLMs are listed and discussed. Table 6 gives the unique features of these models and the prompt techniques used in the selected models. Also, how these prompt techniques are used in the models is presented in detail.

4.3.2. Retrieval augmented generation

Although IFMs have shown impressive performance across various industrial tasks, they are not perfect due to limitations in the accuracy and availability of training data. As a result, IFMs may occasionally generate output that seems reasonable but is not logical or correct, named "hallucination" [207]. RAG [208] is designed to integrate industry-specific data and knowledge bases to enhance the accuracy and consistency of IFMs without modifying its structure [127]. The RAG technology contains the external database, retrieval algorithm, and generation algorithm [209].

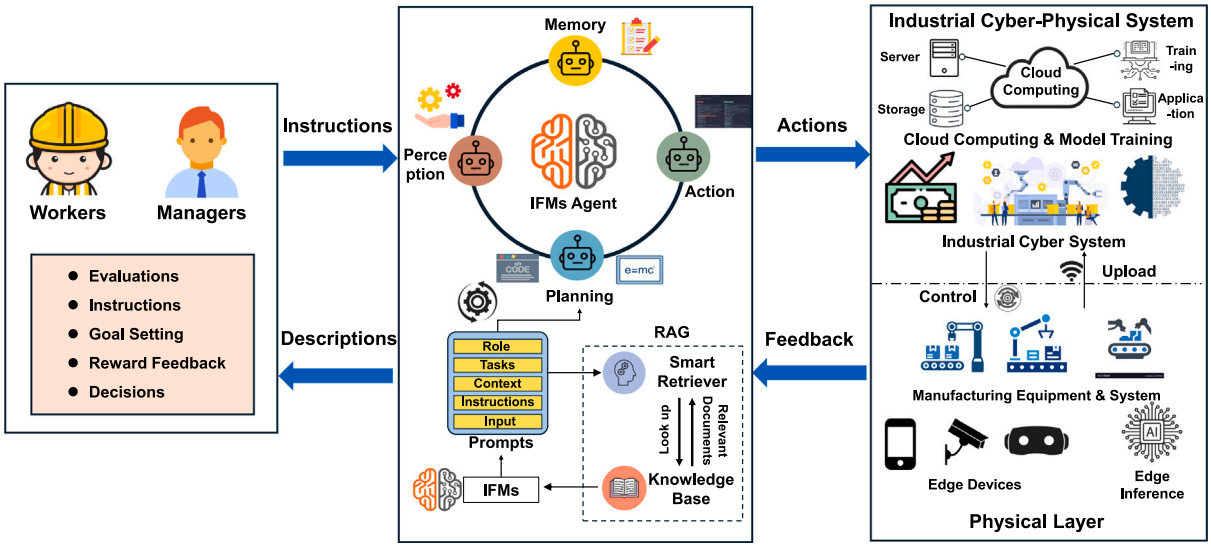

Fig. 7. Application framework of IFMs for intelligent manufacturing.

Table 5
Summary of the prompting techniques [201,204–206].

Technique	Descriptions	Advantages	Differences
Zero-shot prompting	Instructs the model to perform a task without examples, relying solely on the task description.	Simple to use; requires minimal input preparation.	No examples are provided, purely description-based.
Few-shot prompting	Provides a few examples in the prompt to guide the model.	Helps generalise better for unfamiliar tasks; improves performance for specific scenarios.	Includes examples for guidance, unlike zero-shot prompting.
Cot prompts	Encourages generating intermediate reasoning steps before the final answer.	Improves reasoning and problem-solving; mitigates errors in multi-step problems.	Focuses on step-by-step reasoning, which is absent in other techniques.
Self-consistent	Generates multiple reasoning paths and selects the most consistent answer.	Enhances reliability and reduces variance in outputs.	Aggregates multiple outputs for consistency, unlike single-response techniques.
Generated knowledge prompting	Asks the model to generate relevant background information before solving the task.	Improves performance on tasks requiring domain-specific or contextual knowledge.	Combines knowledge generation with task solving, adding an initial step.
Prompt chaining	Breaks down complex tasks into simpler subtasks, chaining responses sequentially.	Makes complex tasks manageable; reduces errors by tackling subtasks individually.	Divides tasks into multiple prompts instead of handling them as one large task.
Tree of thoughts	Explores multiple reasoning branches before concluding.	Provides diverse solutions and enhances robustness for open-ended tasks.	Explores reasoning alternatives, unlike linear chain-of-thought prompting.
Directional stimulus prompting	Guides responses using specific keywords or cues in the prompt.	Offers precise control over model outputs; ensures responses align with specific goals.	Requires careful crafting of cues for steering outputs.
Program-aided language models (PAL)	Use programming logic to perform complex computations or data manipulation.	Handles complex logic or numerical tasks better than pure natural language reasoning.	Integrates computational logic, unlike purely language-based techniques.
Graph prompting	Uses graph structures to represent relationships and dependencies.	Ideal for tasks involving structured data, like knowledge representation and reasoning.	Employs graph-based reasoning, which is absent in most text-based techniques.

First, external databases without specific public industrial information and knowledge are constructed [210]. The industrial knowledge base typically includes equipment manuals, engineering standards, process parameters, operation logs, production schedules, and fault cases [211,212]. These unstructured texts (e.g., PDFs, text files) are converted into structured formats, such as vector databases, and stored using tools like FAISS, Weaviate, and others [213]. This process ensures efficient retrieval and utilisation of knowledge for various industrial applications.

Then, efficient retrieval algorithms such as BM25 and Dense Retrieval [214,215] fetch the most relevant information from the external database. The retrieval approach cuts documents from an external database and converts them into vectors. When a user inputs a query, the most relevant documents will be searched for from vectors based on their similarity. They are used as contextual information and fed into the IFMs along with the input query [216]. The latest and most relevant domain-specific knowledge can be introduced to IFMs using RAG and deep learning approaches to ensure they can generate precise and reliable outputs.

Finally, the retrieved information and knowledge will be combined with the original query and input into the IFMs. Besides, the LangChain framework [217] will guide and constrain the inference process of IFMs. As a result, the IFMs will be able to generate much more reliable outputs consistent with logic. For instance, operators can

Table 6
Summary of the use of prompt methods in commonly used LLMs [206].

Model	Special features	Prompt techniques	Details on usage
ChatGPT	General-purpose, conversational AI.	Zero-shot, Few-shot, Chain-of-Thought	Zero-shot: ChatGPT directly follows instructions in natural language. Relies on pre-training data for accurate answers without prior examples. Few-shot: Uses in-context examples to adjust its predictions for specific tasks (e.g., classification). CoT: Generates step-by-step explanations or reasoning paths, helpful for tasks like math or logic.
GPT-4	Advanced reasoning, multilingual capabilities	Advanced reasoning, Multilingual prompting	Advanced reasoning: Handles complex queries requiring multi-step logical deductions by interpreting the relationships between concepts. Multilingual prompting: Process prompts in various languages, retaining high fidelity in language translation and generation tasks.
Claude 3	Long context handling (200K tokens)	Long document summarisation, Enhanced recall	- Long document summarisation: Leverages its extensive context window to summarise lengthy texts without losing critical information. Enhanced recall: Retains high accuracy for large prompts, ensuring no loss of details or context. This makes it ideal for document analysis or academic reviews.
LLaMA	Adaptable to fine-tuning	Instruction following	Instruction: Models in the LLaMA family are fine-tuned to align with task-specific instructions, adapting to domain-specific tasks efficiently.
Flan	Fine-tuned for task generalisation	Task decomposition, Multitasking	Task decomposition: Automatically breaks complex instructions into smaller tasks for sequential execution. Multitasking: Capable of simultaneously executing multiple interrelated tasks, such as generating and analysing text.
Gemini	Handles complex and multitask prompts	Advanced task chaining	Advanced task chaining: Can manage multiple stages of tasks in a single prompt by retaining focus on task dependencies, ensuring coherent outputs for complex workflows.
Code Llama	Specialised for code generation	Code-specific prompts	Code-specific prompts: Optimised for programming tasks, it can handle prompts tailored for generating, debugging, or optimising code in various languages. Understands programming context and syntax.
Mistral	High-performance text generation	Creative writing prompts	Creative writing prompts: Uses its high-quality text generation capabilities to produce imaginative, fluent, and stylistically consistent outputs for storytelling or ideation tasks.
Phi-2	Excels in mathematical reasoning	Step-by-step problem-solving	Step-by-step problem-solving: Processes math queries by explaining intermediate steps, reducing errors in tasks requiring precision and logical progression.
Sora	Proficient in translation	Multilingual translation	Multilingual translation: Designed to translate between languages efficiently by utilising large multilingual datasets, preserving nuances and tone in the translated text.
OLMo	Open-ended content generation	Open-ended prompts	Open-ended prompts: Ideal for creative or exploratory tasks with no fixed answers, allowing for diverse outputs such as essays, opinion pieces, or brainstorming ideas.
Grok-1	Knowledge retrieval	Fact-based question-answering	Fact-based question-answering: Retrieves factual information accurately by synthesising relevant data from training knowledge or external sources.
Mixtral	Multimodal integration	Multimodal task execution	Multimodal task execution: Processes text-based prompts and other data modalities like images, enhancing its ability to deliver integrated solutions across data formats.

follow a simplified workflow in equipment maintenance scenarios to identify the causes of equipment failures [218]. When a production issue arises, RAG can assist IFMs in searching for troubleshooting guides and manuals to identify potential causes.

4.3.3. Agent engineering

IFMs are capable of analysing and optimising manufacturing processes. However, their full potential for industrial tasks can only be realised when integrated with manufacturing systems [219]. Therefore, integrating IFMs with intelligent manufacturing systems is crucial to solving complex industrial tasks effectively [220]. Agent engineering can be utilised to develop the human-agent-cyber–physical system (HACPS) with capabilities in perception, reasoning, decision-making, and optimisation [221] to achieve higher levels of intelligent manufacturing.

Agents can autonomously perform tasks based on their perception of the environment and defined goals [222]. The agent mediates between different subsystems in the manufacturing system, enabling distributed decision-making and control [223]. It allows the manufacturing system to understand human instructions, dynamically plan complex tasks, invoke different IFMs to execute tasks, and continuously learn to improve performance [224]. The Agent comprises four key components: perception, memory, planning, and action. Table 7 summarises some key technologies for these four modules.

The perception module utilises auxiliary tools such as API calls and plugin extensions to perceive the environment and execute decisions [225]. For example, ChatGPT can be utilised to understand human instructions, and ChatPDF can be used to parse documents and to realise text-to-image generation.

The planning module is responsible for the thinking and decision-making of the Agent [226]. It is able to decompose complex tasks into executable subtasks and evaluate execution strategies [227]. The ReAct and CoT [228] technologies can be utilised to help Agents decompose industrial tasks as several subtasks and solve them step by step.

Memory module [229] refers to the storage and recall of information, which encompasses both short-term and long-term memory. Short-term memory is used to store conversational context for multi-turn interactions. Long-term memory is used to store industrial details, characteristics, and more. The memory information is stored in vector formats for fast retrieval. The memory and planning modules are always deployed in the cloud platform to perform complex computations and inferences [230].

Action module [231] can carry out specific actions based on decisions and interact directly with the manufacturing system. It can

Table 7
Summary of agent engineering of LLMs.

Module	Technologies	Details
Perception module	Natural language perception	They focus on processing and understanding pure text data, with capabilities such as text comprehension, generative writing, and text translation. They are commonly used in tasks like intelligent dialogue systems, document translation, and document generation.
	Visual perception	These models are designed to handle visual data, offering image classification and scene segmentation capabilities. They are widely applied in image retrieval, visual question answering, and multimodal generation tasks.
	Audio and Video perception	They process temporal features in audio and video data, enabling functions such as speech-to-text conversion and speech and video recognition. Common applications include virtual assistants, subtitle generation, conference transcription, and video classification.
	Environmental and Contextual perception	They handle contextual information from sensors and the environment, supporting structured and unstructured data processing. They can process diverse input modalities (text, images, videos, and sensor data) and are frequently applied in tasks such as autonomous navigation, robotic perception, 3D data understanding, and augmented reality.
	Multimodal perception	They integrate multiple input modalities, such as text, images, audio, and video, allowing for multimodal content generation and cross-domain understanding of complex tasks. They are often used in tasks like image-text question answering and multimodal dialogues.
Planning module	Task decomposition approach	Task decomposition approach breaks down a task into several simpler subtasks and creates a step-by-step plan to execute them. HuggingGPT [233] utilises various multimodal models from HuggingFace to build intelligent agents, decomposes tasks provided by humans, selects appropriate models as dependencies between subtasks, and generates the final response. Chain-of-Thought [234] leverages prior knowledge to construct reasoning trajectories, guiding the LLMs to tackle complex problems by utilising its reasoning capabilities for task decomposition. ReAct [235] decouples reasoning and planning, alternates between decomposing and planning
	Multi-plan selection	It combines multi-plan generation and optimal plan selection to facilitate task planning. The self-consistency method [236] generates several distinct reasoning paths using a sampling strategy during the decoding process. The optimal plan is determined using a simple majority voting strategy, where the plan with the most votes is selected. The Tree of Thoughts [237] introduces two strategies for generating plans: sampling and proposing. There are responsible sampling plans during decoding. The proposed strategy explicitly instructs the LLMs to generate diverse plans.
	The external planner	Its approach combines LLMs with external specialised planners for task planning, including symbolic and neural planners. Symbolic Planners integrate LLMs with symbolic reasoning to identify the optimal path to reach the target state [238]. Neural planners are deep models trained on collected planning data using reinforcement learning or imitation learning [239]. These models exhibit effective planning capabilities within specific domains.
Memory module	RAG-based memory	RAG technology stores memory in storage media and retrieves memories based on their relevance to the current context. Historical memories are stored in the form of tables, text formats, and encoded vectors. REMEMBER stores historical memories in the form of Q-value tables [240], where each record is represented as a tuple (environment, task, action, Q-value). MemoryBank [241] and RecMind [242] use text encoding models to encode each memory into a vector and establish an indexing structure. Generative Agents [243] store the daily experiences in text formats. During retrieval, the description of the current state is used as a query to retrieve relevant memories from the memory pool.
	Embodied memory	Embodied memory leverages real-world data to fine-tune IFMs, embedding memory into model parameters as a form of long-term memory [244]. This data typically includes interaction records between agents and the environment (historical data of successes and failures), domain-specific common-sense knowledge, and task-related prior knowledge. Memory embedding is achieved using model fine-tuning methods that train only a small subset of parameters, such as Low-Rank Adaptation (LoRA), QLoRA, and P-tuning [245].
Action module	Integrated with the system	The agent module executes actions by interfacing with internal capabilities and manufacturing systems. It can produce text, images, or other outputs directly from the LLMs. In addition, it also invokes APIs, databases, or external software for task execution. To achieve physical control, it is integrated with hardware, controlling physical systems or robots.

achieve efficient interaction and real-time control of devices, software, and platforms within the intelligent manufacturing system through communication protocols such as Socket and Modbus. The action module is often deployed on edge computing devices that can interact directly with local equipment to improve real-time performance [232].

The IFMs agent, as well as human and intelligent manufacturing systems, constitute the HACPS. Humans include roles such as operators, managers, and after-sales services. Humans can set decision-making goals based on specific task requirements during the interaction between humans and agents. Agents will actively decompose tasks, plan task workflows, optimise objectives, and design various solutions [246]. Humans can also provide reward-punishment feedback mechanisms to Agents, enabling them to optimise their strategies iteratively [247]. Agent leverages their perception capabilities to perceive and analyse industrial environments and further utilise their optimisation and decision-making abilities to design and generate corresponding instructions or outputs. Finally, the agent will interact with the manufacturing system to execute specific tasks, such as defect detection and safety monitoring [248].

5. Industrial applications of IFMs

With six core capabilities—including question answering, scene cognition, scientific discovery, process decision-making, terminal control, and content generation, IFMs can be applied throughout the entire manufacturing lifecycle, which includes research and design, manufacturing, optimisation and management, as well as maintenance [49]. The core functions and typical application scenarios of IFMs in the

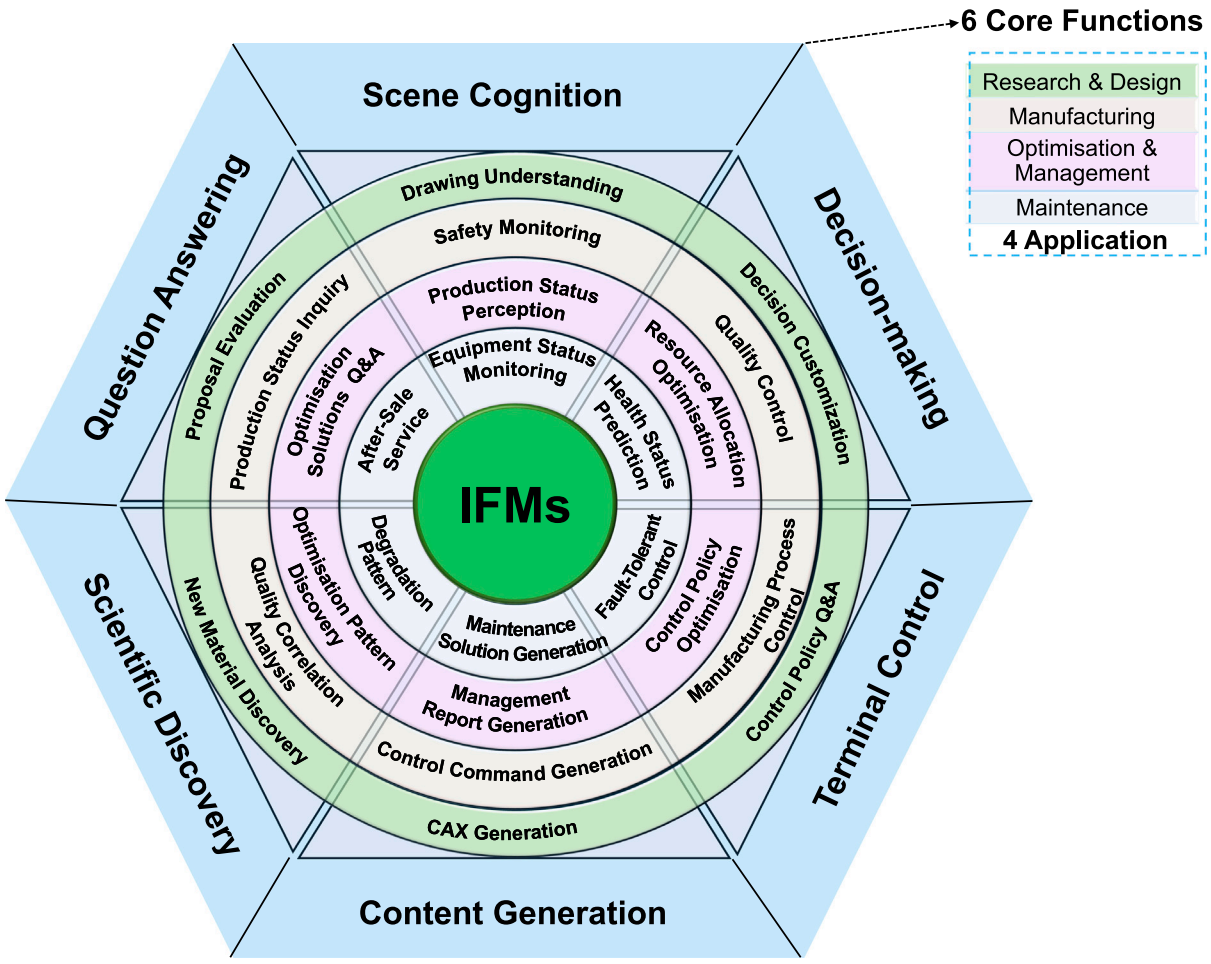

Fig. 8. Core functions and typical application scenarios of IFMs.

manufacturing life cycle are illustrated in Fig. 8. Some existing research on IFMs applications in different industrial tasks is also discussed to highlight their significant impact on intelligent manufacturing.

5.1. Research & design

IFMs are applied in product research & design to support advanced simulations and rapid prototyping. Product research & design heavily rely on human experience, intuition, and skill. However, human expression is always ambiguous. This can lead to misunderstandings and communication barriers during discussions of design concepts. IFMs not only possess the capability to understand the natural language but can also infer the underlying beliefs, emotions, desires, intentions and psychological states behind the language [249,250]. The IFMs collaborative design mechanism may become the core technology of human-centred intelligent manufacturing in the stage of Industry 5.0 [251,252].

Xu [104] proposed an IFMs-augmented multimodal collaborative design (IFMs-MCD) framework. It consists of three stages: machine-to-machine (M2M) collaboration, human-to-machine (H2M) collaboration, and human-to-human (H2H) collaboration. In the M2M collaboration stage, IFMs create a multi-agent iterative mechanism. This mechanism integrates key information from stakeholders, including consumers, merchants, designers, and engineers. It converts design discussions into design semantics during the iteration process. Next, it generates and evaluates design proposals based on these semantics. It effectively evaluates various aspects of the conceptual product, such as functionality, appearance, materials, structure, and pricing, while summarising and unifying semantic relationships. Finally, it utilises specific modules of IFMs to perform quantitative evaluation, ensuring a comprehensive design assessment.

In the H2M collaboration stage, generative AI tools and mixed reality (MR) technologies enable participants to engage in a more intuitive and immersive design environment. This stage consists of two steps: prototype development review and prototype testing review. First, participants evaluate and optimise the 2D and 3D design proposals generated by IFMs to enhance their usability. Next, the designs are materialised and projected onto the physical models as 3D visualisations via MR, allowing for a final review of appearance and functionality.

The H2H collaboration stage encompasses the entire collaborative design process, with all stakeholders actively participating in discussions. These discussions focus on product goals, requirements, quality control, testing, and feedback. Throughout the iteration, evaluation, and revision of design proposals, all stakeholders play a vital role in ensuring that the outcomes align with collective expectations.

The applications of IFMs in product design can effectively address the inherent complexities and communication challenges in collaborative design. It enhances the efficiency of human-machine interactions while reducing costs. It can provide users with a more innovative, real-time, and highly realistic collaborative platform, thus significantly improving design efficiency and lowering costs.

5.2. Human–robot collaboration

The rapid development of IFMs has propelled HRC to a new phase [253]. They are able to achieve smoother and more efficient HRC [254,255] by understanding complex natural language and inferring human intentions.

Wang [256] proposed a navigation method for collaborative robots (cobots) that integrates LLMs and LVMs for visual and language-driven collaboration. The LLMs and LVMs are used to achieve 3D scene reconstruction and annotation at first. Workstation operators can utilise natural language commands to instruct the cobot to navigate to any location and scan the environment to generate 3D point cloud models. These 3D models are stored in PLY format and subsequently annotated to extract scene semantics and labelled scene information.

Then, LLMs are utilised to interpret natural language commands and translate them into executable code. Using a fine-tuning approach, the IFM is trained to learn contextual information and convert natural language commands into executable Python code to guide cobot navigation. The IFM is always optimised by learning multiple user-provided navigation tasks, enabling iterative code updates and refinements.

Finally, the fine-tuned IFM is utilised to generate a navigation grid by incorporating the pathfinder path planning algorithm. Direction vectors between adjacent path points are calculated and converted into quaternions to guide the robot's movements. This process is iteratively repeated to control the direction of cobots, execute the required operations, and achieve the necessary targets.

With the help of IFM, the system can successfully identify targets from language instructions and plan paths for the cobot to reach designated locations. Even in complex scenarios where instructions include multiple sub-goals and distractors, it can also accurately parse instructions, generate corresponding code, and guide the cobot to complete all sub-goals in the correct order. This demonstrates that IFMs can enable cobots to execute navigation tasks effectively in HRC scenarios.

In addition, the applications of the IFMs on intelligent process planning and smart assembly have been actively explored, given their common-sense knowledge and reasoning capabilities derived from foundation models [37,257,258]. As an example, an adaptive process management approach leveraging LLMs was introduced to support precise manufacturing process planning. It utilises LLMs to convert user-provided instructions into structured task sequences, thereby improving the agility and adaptability of production systems. Starting from informal user input, the system performs a formal refinement step, followed by the generation of detailed procedural steps. These components are then composed into a coherent workflow. Finally, state machines are employed to verify both the logical consistency and operational safety of the resulting processes [257,259]. In industrial robotics scenarios such as automated assembly or disassembly, certain operations (e.g., insertion) require heightened precision and involve complex dynamics, including forceful contact, friction response, and delicate actuation [260,261]. Designing a universal policy to handle such tasks remains difficult, especially when precision demands integration of rich sensory feedback like force or torque signals [262]. To address this, an LLM-driven global policy was proposed to dynamically delegate control to a predefined set of specialised skills, each trained to execute fine-grained, high-accuracy tasks through contextual switching [263].

5.3. Operations & management

IFMs can leverage their ability to optimise the scheduling of production plans. As manufacturing evolves towards multi-type and small-batch production, flexible production plays a vital role in the manufacturing system [264]. With strong decision-making abilities, IFMs can effectively improve the scheduling efficiency of manufacturing systems [265].

Huang [63] innovatively combines LLMs with evolutionary algorithms (such as genetic algorithms) to develop an automatic programming method called Population Self-Evolution (SeEvo), which is capable of generating adaptive scheduling rules and production planning. The method consists of an LLM and the SeEvo framework, which can deeply understand tasks, machines, and scheduling constraints, so as to automatically generate and optimise scheduling strategies based on real-time dynamic environmental changes. In this framework, LLMs are used to automatically generate scheduling rules by comprehensively understanding tasks, machines, and scheduling constraints. The SeEvo framework adopts a "self-evolution" strategy to continuously evolve through the simulation of multiple individuals within a population to optimise scheduling strategies. By combining the powerful code generation capabilities of LLMs with the evolutionary mechanism of SeEvo, the method generates flexible and efficient scheduling algorithms. The experimental results demonstrate that SeEvo outperforms traditional approaches such as genetic programming (GP), gene expression programming (GEP), and deep reinforcement learning (DRL) methods in unseen and dynamic scenarios, making it a promising approach for real-world manufacturing environments. The experimental results on DMU dataset are shown in Table 8.

In Table 8, the SeEvo(GLM3) and SeEvo(GPT3.5) are developed based on the GLM3 and GPT3.5. The UB is the best-known solution for the dataset. The comparative experimental results on DMU Datasets show that the SeEvo method achieves better scheduling results compared to other methods, which proves the effectiveness of LLMS in scheduling optimisation.

5.4. Quality control

IFMs can also be adopted in industrial vision tasks to ensure product quality [266]. Although IFMs perform well in general recognition tasks, they often lack sufficient knowledge in specialised industrial scenarios, particularly in adapting to specific industrial images and understanding domain-specific terminology [267].

To tackle the issues above, Wang proposes a two-stage adaptation strategy for industrial vision tasks named DefectGLM [268]. The strategy comprises three key components: a visual encoder, a query transformer, and a language encoder. The visual encoder is responsible for processing visual information. The query transformer bridges the gap between multimodal information. The language encoder manages and generates language-based outputs.

In the first adaptation stage, LoRA is employed to enable DefectGLM to adapt to specific industrial image patterns, thereby enhancing its image recognition capabilities. To minimise training costs and prevent catastrophic forgetting, all pre-trained parameters of the visual encoder and language decoder are frozen, with only the LoRA parameters updated. This approach allows DefectGLM to rapidly adapt to new task requirements while preserving its pre-trained knowledge. Additionally, contrastive representation learning is utilised to achieve self-supervised adaptation in the pre-trained visual encoder, enabling it to capture transferable representations of defect patterns more effectively.

In the second adaptation stage, DefectGLM employs a vision-language instruction fine-tuning method to align the model with domain-specific knowledge. A generated image-text dataset is used as supervisory signals to train the model to provide accurate diagnostic answers based on industrial image inputs and instructional prompts. During this process, the visual encoder extracts features from defect patterns, the query transformer maps these features into the language embedding space, and the language decoder generates natural language responses.

The DefectGLM is evaluated on a large-scale multimodal wafer defect dataset with 36 types of defects covering single defect patterns and mixed-type defect patterns. Experimental results demonstrate that DefectGLM achieves outstanding performance in wafer defect recognition, with an accuracy exceeding 99%, significantly outperforming classical visual methods such as VGG16 [269], ResNet [270], and DenseNet [271]. The visual question answering (VQA) results of DefectGLM in the wafer defect dataset are shown in Fig. 9. It can be observed that DefectGLM can follow different prompt instructions and produce clear and concise responses, which enables more efficient human-machine interaction between the operator and the system.

Table 8
Experiment results on DMU benchmark [63].

Cases	Size	Random	DRL	GP	GEP	SeEvo(GLM3)	SeEvo(GPT3.5)	UB
DMU03	20 × 15	3827	3303	3540	3651	3462	**3238**	2731
DMU08	20 × 20	4228	4098	3802	4023	**3728**	**3728**	3188
DMU13	30 × 15	5451	4708	4765	4812	**4658**	4709	3681
DMU18	30 × 20	5326	4800	**4696**	4917	4724	4724	3844
DMU23	40 × 15	5948	5240	5391	5595	**5151**	5258	4668
DMU28	40 × 20	6737	5948	6017	6142	**5838**	5944	4692
DMU33	50 × 15	6890	6458	6109	6081	**6029**	**6029**	5728
DMU38	50 × 20	7685	7275	7267	7501	**7168**	7170	5713

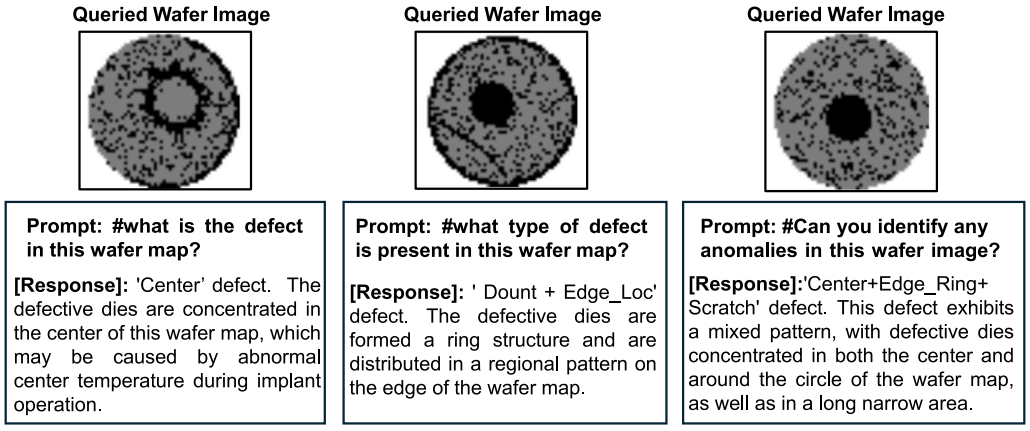

Fig. 9. Detection examples of the DefectGLM.

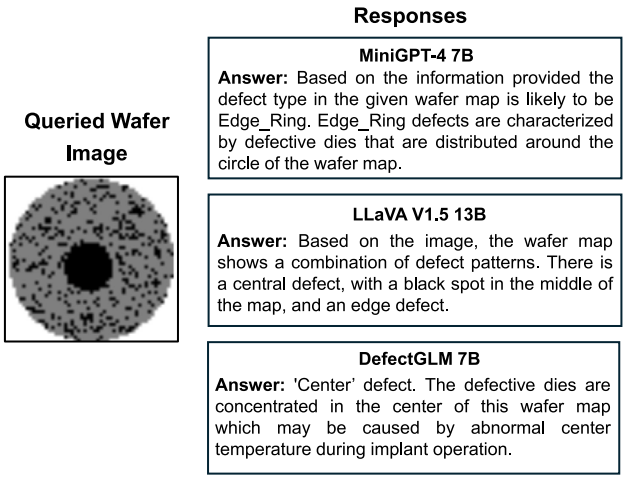

Fig. 10. Comparative experiments with other LLMs.

In addition, DefectGLM is also compared with GPT-4 [74] and LLaVA [272]. Their answers are evaluated in terms of diagnostic correctness and pattern understanding on a wafer with a centre defect, results are shown in Fig. 10. It can be seen that GPT-4 provides incorrect diagnoses of wafer defect patterns. The LLaVA correctly identifies the central black spot in the wafer map but mistakenly identifies edge defects that do not exist. In contrast, DefectGLM provides accurate diagnoses and pattern descriptions, while also providing the causes of the defect. This demonstrates that the proposed two-stage fine-tuning strategy can effectively improve the performance of LLMs by learning from specific industrial scenarios.

6. Key challenges of IFMs

6.1. Challenges in data-level

(1) Low Quality Industrial Data

Industrial data collected during production processes primarily consists of equipment status and periodic production data, most of which is repetitive with limited critical information. This leads to industrial data exhibiting low-quality and low-density characteristics, significantly diminishing its value for model training and applications. To tackle these challenges, domain-knowledge-enhanced (operating rules, process parameters) data cleaning, analysis, and integration methodologies should be developed to effectively extract critical features from industrial data. Additionally, dynamic data modelling approaches should be designed to adaptively adjust data processing strategies in real-time to changing operational conditions.

(2) Heterogeneous Industrial Data

Industrial production is a complex process involving several stages, with factors such as data sources, data collection methods, and varying timestamps leading to heterogeneous data structures. This heterogeneity poses significant challenges for the effective management and utilisation of industrial data, particularly for the training and application of IFMs. To overcome this problem, core manufacturing architectures that consist of multimodal data processing, cross-temporal data alignment, and data fusion approaches should be developed to enable efficient management and processing of heterogeneous industrial data.

(3) Industrial Data Privacy

Industrial data always involves proprietary process parameters and trade secrets of enterprises. The leakage of such key information may directly impact the competitiveness of enterprises. Unlike personal data, industrial data is generally difficult to anonymise as specific production processes or parameters can often be easily reverse-engineered. Some approaches should be designed to ensure the privacy of industrial data. For instance, federated learning can be used to train models

locally and then securely aggregate parameters without sharing raw data. Differential privacy algorithms can also be studied to introduce noise to training data, thus preventing reverse-engineering of original data from model parameters.

6.2. Challenges in model-level

(1) Trustworthy IFMs

The industrial production environment typically involves complex processes, high-precision operation controls, and strict safety standards. Decision errors can directly result in accidents and significant economic losses. However, IFMs may occasionally generate results that are inconsistent with real-world conditions, which severely limits the applications of IFMs in industrial scenarios. This problem may stem from the noise in the training data, the bias of input samples, or the overfitting caused by the model complexity. Therefore, XAI approaches, including causal inference, Grad-CAM, and neural-symbolic networks, should be studied to ensure that IFMs can deliver reliable performance across diverse industrial environments.

(2) Real-time Performance

Industrial production requires exceptionally high real-time responsiveness, with some scenarios demanding responses within milliseconds or even microseconds. Due to constraints in computational resources and real-time capabilities, the size and complexity of IFMs must be maintained within a reasonable range. To enable real-time inference, IFMs must deliver exceptional performance while minimising computational complexity and memory usage. Model lightweight techniques such as pruning, quantisation, and knowledge distillation can be explored to reduce the computational requirements while maintaining model performance.

(3) Integration with Knowledge

Industrial tasks require IFMs to possess a deep understanding of domain-specific knowledge. While IFMs demonstrate powerful generalisation capabilities, they often struggle to fully capture the unique characteristics of specific domains [273]. This limitation primarily stems from a lack of specialised industrial knowledge during the training and inference, which strictly restricts their performance on specialised tasks. To address this challenge, knowledge graphs and graph neural networks can be used to achieve the domain knowledge embedding of IFMs and enable them to deliver precise solutions that meet the complex and diverse demands of industrial tasks.

6.3. Challenges in application-level

(1)Integration with Manufacturing System

Different industrial domains exhibit varied business logic, production processes, and standards. For instance, the automotive manufacturing, electronics, and chemical production industries have unique production requirements, leading to highly customised configurations of manufacturing execution systems (MES) or enterprise resource planning (ERP) systems. The integration methods between IFMs and manufacturing systems need to be explored to ensure that they can meet the unique characteristics and requirements of various industrial sectors while maintaining the efficiency of data management and industrial production.

(2) Maintenance Cost

Maintaining and updating IFMs is an ongoing endeavour that may require substantial cost. As industrial environments and data continuously evolve, regular retraining and fine-tuning are required to sustain model performance. Therefore, a comprehensive model management system and real-time monitoring mechanisms must be established to dynamically update IFMs, thereby ensuring their stability and reliability in practical applications. Besides, it is also necessary to design intelligent model updating methods (e.g., incremental learning approaches) to reduce maintenance costs and training time, making them suitable for dynamic industrial environments.

(3) Deployment Cost

The deployment of IFMs is also a significant challenge. In industrial applications, IFMs often need to be deployed locally to ensure data security, especially in scenarios involving sensitive information, such as core parameters in manufacturing processes and patient privacy data in healthcare. However, the development, training, and updating of large models typically require substantial computational resources, including high-performance servers, storage devices, and network infrastructure, which are usually managed in the cloud. Therefore, it is essential to explore collaboration mechanisms between edge and cloud computing to allocate computational resources effectively, ensuring the secure deployment and update of large models while maintaining real-time performance and scalability.

7. Conclusions and future directions

The emergence of IFMs marks a transformative shift in intelligent manufacturing. This paper systematically analyses the development of IFMs from the data level, model level, and application level, highlighting their potential to revolutionise manufacturing. By integrating multimodal industrial data, domain-specific knowledge, and artificial intelligence technologies, IFMs demonstrate profound capabilities in scene understanding, content generation, and decision-making. IFMs have also been successfully implemented in various industries, including robotics, steel production, energy, and construction, showcasing their scalability and adaptability in industrial scenarios. IFMs also face significant challenges regarding multimodal data, model trustworthiness, and integration with manufacturing systems [274]. Addressing these challenges requires coordinated efforts from academia and industry to refine technologies and promote their applications.

In terms of future work, several key directions can be pursued to advance IFMs further:

- Cross-domain knowledge integration: The future of IFMs lies in their ability to seamlessly incorporate cross-domain knowledge, such as combining manufacturing data with supply chain, laws, or market analytics. This integration could further optimise production processes and decision-making.
- Improving model robustness and explainability: While IFMs demonstrate strong performance, their black-box nature still restricts their application in industrial scenarios. Improving the transparency of its decision-making process will help ensure wider acceptance of IFMs in safety–critical applications. Technologies such as XAI should be explored to make it more trustworthy.
- Collaboration between small and large models: Future research should explore the collaboration mechanism between small models and large models. Small models deployed on edge devices can handle time-sensitive tasks with fast inference. In contrast, large models in the cloud can focus on more computationally intensive processes, such as deep analysis and process optimisation. This hierarchical collaboration will ensure real-time performance while maintaining accuracy and scalability.
- Advancing human-agent collaboration: Integrating IFMs with intelligent manufacturing systems should be expanded to enable smoother and more efficient human-agent collaboration. Agent engineering and advanced manufacturing systems can ensure iterative learning and continuous improvement.
- Ethics and sustainability: Future research must address the ethical and environmental considerations of IFMs. This includes ensuring data privacy, reducing computational energy consumption, and aligning IFMs with sustainable manufacturing practices.

Addressing these areas will enable IFMs to overcome current limitations and unlock unprecedented opportunities for innovation in intelligent manufacturing. As industries transition towards Industry 5.0, IFMs are poised to become integral to realising human-centred, sustainable, and efficient manufacturing systems.

CRediT authorship contribution statement

Shuxuan Zhao: Writing – original draft, Methodology, Investigation, Conceptualization. **Sichao Liu:** Writing – original draft, Investigation. **Yishuo Jiang:** Writing – original draft, Investigation. **Bo Zhao:** Writing – original draft. **Youlong Lv:** Supervision, Funding acquisition. **Jie Zhang:** Supervision, Funding acquisition. **Lihui Wang:** Supervision. **Ray Y. Zhong:** Supervision, Investigation, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

This research is supported by Guangdong Special Support Talent Program—Innovation and Entrepreneurship Leading Team (2019BT02S593), RGC Research Impact Fund (R7036-22), RGC Theme-based Research Scheme (T32-707-22-N), RGC GRF project (17202124), Innovation and Technology Fund (ITF), Hong Kong (ITT/024/24LP) and Public Policy Research Funding Scheme (PPRFS) (2024.A8.154.24D), the National Natural Science Foundation of China (No. 52375485; No. U24A20262), Shanghai Natural Science Foundation, China (No. 22ZR1403000), the Swedish Research Council (Vetenskapsrådet) under award 2023–00493. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

[1] Zhong RY, Ge W. Internet of things enabled manufacturing: A review. Int J Agil Syst Manag 2018;11(2):126–54.
[2] Gao L, Shen W, Li X. New trends in intelligent manufacturing. Engineering 2019;5(4):619–20.
[3] Zhang C, Gao Q, Basin MV, Lü J, Liu H. Robust control of multi-line re-entrant manufacturing plants via stochastic continuum models. IEEE Trans Autom Sci Eng 2024;21(4):4923–35.
[4] Ahmed I, Jeon G, Piccialli F. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. IEEE Trans Ind Inform 2022;18(8):5031–42.
[5] da Silva ER, Shinohara AC, Nielsen CP, de Lima EP, Angelis J. Operating digital manufacturing in industry 4.0: The role of advanced manufacturing technologies. Procedia CIRP 2020;93:174–9.
[6] Wang Y, Han Y, Gong D, Li H. A review of intelligent optimization for group scheduling problems in cellular manufacturing. Front Eng Manag 2023;10(3):406–26.
[7] Ling S, Guo D, Li M, Rong Y, Huang GQ. Heterogeneous demand–capacity synchronization for smart assembly cell line based on artificial intelligence-enabled IIoT. J Intell Manuf 2024;35(2):539–54.
[8] Pivoto DGS, de Almeida LFF, da Rosa Righi R, Rodrigues JJPC, Lugli AB, Alberti AM. Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. J Manuf Syst 2021;58:176–92.
[9] Ren J, Cheng Y, Zhang Y, Tao F. A digital twin-enhanced collaborative maintenance paradigm for aero-engine fleet. Front Eng Manag 2024;11(2):356–61.
[10] Tao F, Zhang H, Liu A, Nee AYC. Digital twin in industry: state-of-the-art. IEEE Trans Ind Inform 2019;15(4):2405–15.
[11] Liu S, Wang XV, Wang L. Digital twin-enabled advance execution for human-robot collaborative assembly. CIRP annals 2022;71(1):25–8.
[12] Zheng P, wang H, Sang Z, Zhong RY, Liu Y, Liu C, Mubarok K, Yu S, Xu X. Smart manufacturing systems for industry 4.0: conceptual framework, scenarios, and future perspectives. Front Mech Eng 2018;13(2):137–50.
[13] Yang T, Yi X, Lu S, Johansson KH, Chai T. Intelligent manufacturing for the process industry driven by industrial artificial intelligence. Engineering 2021;7(9):1224–30.
[14] Zhong RY, Xu C, Chen C, Huang GQ. Big data analytics for physical internet-based intelligent manufacturing shop floors. Int J Prod Res 2017;55(9):2610–21.
[15] Wang L. From intelligence science to intelligent manufacturing. Engineering 2019;5(4):615–8.
[16] Nishant R, Kennedy M, Corbett J. Artificial intelligence for sustainability: challenges, opportunities, and a research Agenda. Int J Inf Manage 2020;53:102104.
[17] Kumar SPL. Knowledge-based expert system in manufacturing planning: State-of-the-art review. Int J Prod Res 2019.
[18] Kusiak A, Chen M. Expert systems for planning and scheduling manufacturing systems. European J Oper Res 1988;34(2):113–30.
[19] Zheng P, Xia L, Li C, Li X, Liu B. Towards Self-X cognitive manufacturing network: An industrial knowledge graph-based multi-agent reinforcement learning approach. J Manuf Syst 2021;61:16–26.
[20] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[21] Rai R, Tiwari MK, Ivanov D, Dolgui A. Machine learning in manufacturing and industry 4.0 applications. Int J Prod Res 2021;59(16):4773–8.
[22] Zhao S, Zhong RY, Jiang Y, Besklubova S, Tao J, Yin L. Hierarchical spatial attention-based cross-scale detection network for digital works supervision system (DWSS). Comput Ind Eng 2024;192:110220.
[23] Nie X, Xie G. A novel normalized recurrent neural network for fault diagnosis with noisy labels. J Intell Manuf 2021;32(5):1271–88.
[24] Zonta T, da Costa CA, da Rosa Righi R, de Lima MJ, da Trindade ES, Li GP. Predictive maintenance in the Industry 4.0: A systematic literature review. Comput Ind Eng 2020;150:106889.
[25] Wang S, Lei Y, Lu N, Yang B, Li X, Li N. Graph continual learning network: An incremental intelligent diagnosis method of machines for new fault detection. IEEE Trans Autom Sci Eng 2024.
[26] Li R, Zhuang L, Li Y, Shen C. Intelligent bearing fault diagnosis based on scaled Ramanujan filter banks in noisy environments. IEEE Trans Instrum Meas 2021;70:1–13.
[27] Khan Z, Saghir A, Katona A, Kosztyán ZT. EWMA control chart framework for efficient Maxwell quality characteristic monitoring: An application to the aerospace industry. Comput Ind Eng 2025;200:110753.
[28] Li S, Wang R, Zheng P, Wang L. Towards proactive human–robot collaboration: A foreseeable cognitive manufacturing paradigm. J Manuf Syst 2021;60:547–52.
[29] Zhou K, Liu Z, Qiao Y, Xiang T, Loy CC. Domain generalization: a survey. IEEE Trans Pattern Anal Mach Intell 2023;45(4):4396–415.
[30] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. 2017, arXiv.
[31] Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber-Phys Syst 2023;3:121–54.
[32] Liu Y, Zhang K, Li Y, Yan Z, Gao C, Chen R, Yuan Z, Huang Y, Sun H, Gao J, He L, Sun L. Sora: a review on background, technology, limitations, and opportunities of large vision models. 2024, arXiv:2402.17177.
[33] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, Dollar P, Girshick R. Segment Anything. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 4015–26.
[34] Caruccio L, Cirillo S, Polese G, Solimando G, Sundaramurthy S, Tortora G. Claude 2.0 large language model: Tackling a real-world classification problem with a new iterative prompt engineering approach. Intell Syst Appl 2024;21:200336.
[35] Wang Y, Yang C, Lan S, Fei W, Wang L, Huang GQ, Zhu L. Towards industrial foundation models: framework, key issues and potential applications. In: 2024 27th international conference on computer supported cooperative work in design. CSCWD, 2024, p. 1–6.
[36] Wang H, Wang C, Liu Q, Zhang X, Liu M, Ma Y, Yan F, Shen W. A data and knowledge driven autonomous intelligent manufacturing system for intelligent factories. J Manuf Syst 2024;74:512–26.
[37] Zhang H, Semujju SD, Wang Z, Lv X, Xu K, Wu L, Jia Y, Wu J, Liang W, Zhuang R, et al. Large scale foundation models for intelligent manufacturing applications: a survey. J Intell Manuf 2025;1–52.
[38] Wang J, Xu C, Zhang J, Zhong R. Big data analytics for intelligent manufacturing systems: A review. J Manuf Syst 2022;62:738–52.
[39] Li Y, Zhang Y, Wu J, Xie M. Regularized periodic Gaussian process for nonparametric sparse feature extraction from noisy periodic signals. IEEE Trans Autom Sci Eng 2024.
[40] Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. J Manuf Syst 2018;48:157–69.
[41] Leng J, Zhong Y, Lin Z, Xu K, Mourtzis D, Zhou X, Zheng P, Liu Q, Zhao JL, Shen W. Towards resilience in Industry 5.0: A decentralized autonomous manufacturing paradigm. J Manuf Syst 2023;71:95–114.
[42] Zhao S, Yin L, Zhang J, Wang J, Zhong R. Real-time fabric defect detection based on multi-scale convolutional neural network. IET Collab Intell Manuf 2020;2(4):189–96.
[43] Fakhar Manesh M, Pellegrini MM, Marzi G, Dabic M. Knowledge management in the fourth industrial revolution: mapping the literature and scoping future avenues. IEEE Trans Eng Manage 2021;68(1):289–300.
[44] Xiang W, Yu K, Han F, Fang L, He D, Han Q-L. Advanced manufacturing in industry 5.0: a survey of key enabling technologies and future trends. IEEE Trans Ind Inform 2024;20(2):1055–68.

[45] Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee YT, Li Y, Lundberg S, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. 2023, arXiv preprint arXiv:2303.12712.

[46] Fei N, Lu Z, Gao Y, Yang G, Huo Y, Wen J, Lu H, Song R, Gao X, Xiang T, Sun H, Wen J-R. Towards artificial general intelligence via a multimodal foundation model. Nat Commun 2022;13(1):3094.

[47] Hadi MU, Al Tashi Q, Shah A, Qureshi R, Muneer A, Irfan M, Zafar A, Shaikh MB, Akhtar N, Wu J, et al. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. Authorea Prepr 2024.

[48] Garcia CI, DiBattista MA, Letelier TA, Halloran HD, Camelio JA. Framework for LLM applications in manufacturing. Manuf Lett 2024;41:253–63.

[49] Ren L, Wang H, Dong J, Jia Z, Li S, Wang Y, Laili Y, Huang D, Zhang L, Wu W, Li B. Industrial foundation model: Architecture, key technologies, and typical applications. SCI SIN Inform 2024;54(11):2606.

[50] Saxena S, Prasad S, I M, Shankar A, V V, Gopalakrishnan S, Vaddina V. Automated Tailoring of Large Language Models for Industry-Specific Downstream Tasks. In: Proceedings of the 17th ACM international conference on web search and data mining. WSDM '24, New York, NY, USA: Association for Computing Machinery; 2024, p. 1184–5.

[51] Saka A, Taiwo R, Saka N, Salami BA, Ajayi S, Akande K, Kazemi H. GPT models in construction industry: Opportunities, limitations, and a use case validation. Dev Built Environ 2024;17:100300.

[52] Xu M, Yin W, Cai D, Yi R, Xu D, Wang Q, Wu B, Zhao Y, Yang C, Wang S, Zhang Q, Lu Z, Zhang L, Wang S, Li Y, Liu Y, Jin X, Liu X. A Survey of resource-efficient LLM and multimodal foundation models. 2024, arXiv:2401.08092.

[53] Yin S, Li X, Gao H, Kaynak O. Data-based techniques focused on modern industry: an overview. IEEE Trans Ind Electron 2015;62(1):657–67.

[54] Oztemel E, Gursev S. Literature review of industry 4.0 and related technologies. J Intell Manuf 2020;31(1):127–82.

[55] He B, Chen W, Li F, Yuan X. Directed acyclic graphs-based diagnosis approach using small data sets for sustainability. Comput Ind Eng 2023;176:108944.

[56] Chandrasekhar A, Chan J, Ogoke F, Ajenifujah O, Farimani AB. AMGPT: A large language model for contextual querying in additive manufacturing. 2024, arXiv:2406.00031.

[57] Rosati R, Antonini F, Muralikrishna N, Tonetto F, Mancini A. Improving industrial question answering chatbots with domain-specific LLMs fine-tuning. In: 2024 20th IEEE/ASME international conference on mechatronic and embedded systems and applications. MESA, 2024, p. 1–7.

[58] Tang C, Huang D, Ge W, Liu W, Zhang H. GraspGPT: leveraging semantic knowledge from a large language model for task-oriented grasping. 2023, arXiv:2307.13204.

[59] Liu J, Feng Y, Lu C, Fei C. Knowledge embedding synchronous surrogate modeling for multi-objective operational reliability evaluation of complex mechanical systems. Comput Ind Eng 2024;196:110482.

[60] Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput Surv 2023;55(9):195:1–35.

[61] Arslan M, Mahdjoubi L, Munawar S. Driving sustainable energy transitions with a multi-source RAG-LLM system. Energy Build 2024;324:114827.

[62] Liu Z, Yao W, Zhang J, Yang L, Liu Z, Tan J, Choubey PK, Lan T, Wu J, Wang H, Heinecke S, Xiong C, Savarese S. AgentLite: a lightweight library for building and advancing task-oriented LLM agent system. 2024, arXiv:2402.15538.

[63] Huang J, Li X, Gao L, Liu Q, Teng Y. Automatic programming via large language models with population self-evolution for dynamic job shop scheduling problem. 2024, arXiv:2410.22657.

[64] Fan H, Zhang H, Ma C, Wu T, Fuh JYH, Li B. Enhancing metal additive manufacturing training with the advanced vision language model: A pathway to immersive augmented reality training for non-experts. J Manuf Syst 2024;75:257–69.

[65] Liu S, Zhang J, Yi S, Gao R, Mourtzis D, Wang L. Human-centric systems in smart manufacturing. In: Manufacturing from Industry 4.0 to Industry 5.0. Elsevier; 2024, p. 181–205.

[66] Tinnes C, Ristin M, Hohenstein U, Fathi K, Van De Venn HW. From unstructured product descriptions to structured data for industry 4.0 with ChatGPT. In: 2024 IEEE 7th international conference on industrial cyber-physical systems. ICPS, IEEE; 2024, p. 1–8.

[67] Ma P, Wang T-H, Guo M, Sun Z, Tenenbaum JB, Rus D, Gan C, Matusik W. LLM and simulation as bilevel optimizers: a new paradigm to advance physical scientific discovery. 2024, arXiv:2405.09783.

[68] Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, Min Y, Zhang B, Zhang J, Dong Z, Du Y, Yang C, Chen Y, Chen Z, Jiang J, Ren R, Li Y, Tang X, Liu Z, Liu P, Nie J-Y, Wen J-R. A survey of large language models. 2024, arXiv:2303.18223.

[69] Floridi L, Chiriatti M. GPT-3: its nature, scope, limits, and consequences. Minds Mach 2020;30(4):681–94.

[70] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing 2021;452:48–62.

[71] Zhang C, Zhang Y, Liu S, Wang L. Transfer learning and augmented data-driven parameter prediction for robotic welding. Robotics and Computer-Integrated Manufacturing 2025;95:102992.

[72] Deng H, Khan S, Erkoyuncu JA. An investigation on utilizing large language model for industrial computer-aided design automation. Procedia CIRP 2024;128:221–6.

[73] Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng H-T, Jin A, Bos T, Baker L, Du Y, et al. Lamda: Language models for dialog applications. 2022, arXiv preprint arXiv:2201.08239.

[74] OpenAI, Achiam J, Adler S, Agarwal S, Ahmad. GPT-4 technical report. 2024, arXiv:2303.08774.

[75] Smith S, Patwary M, Norick B, LeGresley P, Rajbhandari S, Casper J, Liu Z, Prabhumoye S, Zerveas G, Korthikanti V, et al. Using deepspeed and megatron to train megatron-turing nlg 530b, a large-scale generative language model. 2022, arXiv preprint arXiv:2201.11990.

[76] Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, Bashlykov N, Batra S, Bhargava P, Bhosale S, Bikel D, Blecher L, Ferrer CC, Chen M, Cucurull G, Esiobu D, Fernandes J, Fu J, Fu W, Fuller B, Gao C, Goswami V, Goyal N, Hartshorn A, Hosseini S, Hou R, Inan H, Kardas M, Kerkez V, Khabsa M, Kloumann I, Korenev A, Koura PS, Lachaux M-A, Lavril T, Lee J, Liskovich D, Lu Y, Mao Y, Martinet X, Mihaylov T, Mishra P, Molybog I, Nie Y, Poulton A, Reizenstein J, Rungta R, Saladi K, Schelten A, Silva R, Smith EM, Subramanian R, Tan XE, Tang B, Taylor R, Williams A, Kuan JX, Xu P, Yan Z, Zarov I, Zhang Y, Fan A, Kambadur M, Narang S, Rodriguez A, Stojnic R, Edunov S, Scialom T. Llama 2: open foundation and fine-tuned chat models. 2023, arXiv:2307.09288.

[77] Bai J, Bai S, Chu Y, Cui Z, Dang K, Deng X, Fan Y, Ge W, Han Y, Huang F, Hui B, Ji L, Li M, Lin J, Lin R, Liu D, Liu G, Lu C, Lu K, Ma J, Men R, Ren X, Ren X, Tan C, Tan S, Tu J, Wang P, Wang S, Wang W, Wu S, Xu B, Xu J, Yang A, Yang H, Yang J, Yang S, Yao Y, Yu B, Yuan H, Yuan Z, Zhang J, Zhang X, Zhang Y, Zhang Z, Zhou C, Zhou J, Zhou X, Zhu T. Qwen technical report. 2023, arXiv:2309.16609.

[78] Sun Y, Wang S, Feng S, Ding S, Pang C, Shang J, Liu J, Chen X, Zhao Y, Lu Y, Liu W, Wu Z, Gong W, Liang J, Shang Z, Sun P, Liu W, Ouyang X, Yu D, Tian H, Wu H, Wang H. ERNIE 3.0: large-scale knowledge enhanced pre-training for language understanding and generation. 2021, arXiv:2107.02137.

[79] Lin J, Men R, Yang A, Zhou C, Ding M, Zhang Y, Wang P, Wang A, Jiang L, Jia X, et al. M6: A chinese multimodal pretrainer. 2021, arXiv preprint arXiv:2103.00823.

[80] Sun X, Chen Y, Huang Y, Xie R, Zhu J, Zhang K, Li S, Yang Z, Han J, Shu X, et al. Hunyuan-large: An open-source moe model with 52 billion activated parameters by tencent. 2024, arXiv preprint arXiv:2411.02265.

[81] Zeng W, Ren X, Su T, Wang H, Liao Y, Wang Z, Jiang X, Yang Z, Wang K, Zhang X, Li C, Gong Z, Yao Y, Huang X, Wang J, Yu J, Guo Q, Yu Y, Zhang Y, Wang J, Tao H, Yan D, Yi Z, Peng F, Jiang F, Zhang H, Deng L, Zhang Y, Lin Z, Zhang C, Zhang S, Guo M, Gu S, Fan G, Wang Y, Jin X, Liu Q, Tian Y. PanGu-α: large-scale autoregressive pretrained Chinese language models with auto-parallel computation. 2021, arXiv:2104.12369.

[82] Zhou C, Li Q, Li C, Yu J, Liu Y, Wang G, Zhang K, Ji C, Yan Q, He L, et al. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. Int J Mach Learn Cybern 2024;1–65.

[83] Jin Y, Li J, Liu Y, Gu T, Wu K, Jiang Z, He M, Zhao B, Tan X, Gan Z, et al. Efficient multimodal large language models: A survey. 2024, arXiv preprint arXiv:2405.10739.

[84] Ren L, Wang H, Dong J, Jia Z, Li S, Wang Y, Laili Y, Huang D, Zhang L, Wu W, Li B. Industrial foundation model: Architecture, key technologies, and typical applications. SCI SIN Inform 2024;54(11):2606.

[85] Ni J, Ábrego GH, Constant N, Ma J, Hall KB, Cer D, Yang Y. Sentence-T5: scalable sentence encoders from pre-trained text-to-text models. 2021, arXiv:2108.08877.

[86] Xue L, Constant N, Roberts A, Kale M, Al-Rfou R, Siddhant A, Barua A, Raffel C. mT5: A massively multilingual pre-trained text-to-text transformer. 2021, arXiv:2010.11934.

[87] Zhou G, Zhang Y, Hu R, Zhang Y. LanYUAN, a GPT large model using curriculum learning and sparse attention. In: Proceedings of the 2023 4th international conference on computing, networks and internet of things. CNIOT '23, New York, NY, USA: Association for Computing Machinery; 2023, p. 265–72.

[88] Alayrac J-B, Donahue J, Luc P, Miech A, Barr I, Hasson Y, Lenc K, Mensch A, Millican K, Reynolds M, Ring R, Rutherford E, Cabi S, Han T, Gong Z, Samangooei S, Monteiro M, Menick JL, Borgeaud S, Brock A, Nematzadeh A, Sharifzadeh S, Bińkowski M, Barreira R, Vinyals O, Zisserman A, Simonyan K. Flamingo: A visual language model for few-shot learning. Adv Neural Inf Process Syst 2022;35:23716–36.

[89] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th international conference on machine learning. PMLR; 2021, p. 8748–63.

[90] Gao M, Bu W, Miao B, Wu Y, Li Y, Li J, Tang S, Wu Q, Zhuang Y, Wang M. Generalist virtual agents: a survey on autonomous agents across digital platforms. 2024, arXiv:2411.10943.

[91] Dang Y, Huang K, Huo J, Yan Y, Huang S, Liu D, Gao M, Zhang J, Qian C, Wang K, Liu Y, Shao J, Xiong H, Hu X. Explainable and interpretable multimodal large language models: a comprehensive survey. 2024, arXiv:2412.02104.
[92] Huang H, Feng Y, Shi C, Xu L, Yu J, Yang S. Free-bloom: Zero-shot text-to-video generator with llm director and ldm animator. Adv Neural Inf Process Syst 2024;36.
[93] Wu D, Li J, Wang B, Zhao H, Xue S, Yang Y, Chang Z, Zhang R, Qian L, Wang B, et al. SparkRA: A retrieval-augmented knowledge service system based on spark large language model. 2024, arXiv preprint arXiv:2408.06574.
[94] Kolter JZ. AlphaCode and "data-driven" programming. Science 2022;378(6624). 1056–1056.
[95] Ye Q, Xu H, Xu G, Ye J, Yan M, Zhou Y, Wang J, Hu A, Shi P, Shi Y, Li C, Xu Y, Chen H, Tian J, Qian Q, Zhang J, Huang F, Zhou J. mPLUG-Owl: modularization empowers large language models with multimodality. 2024, arXiv:2304.14178.
[96] Girdhar R, El-Nouby A, Liu Z, Singh M, Alwala KV, Joulin A, Misra I. ImageBind: One Embedding Space To Bind Them All. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 15180–90.
[97] Wu C, Yin S, Qi W, Wang X, Tang Z, Duan N. Visual ChatGPT: talking, drawing and editing with visual foundation models. 2023, arXiv:2303.04671.
[98] Li J, Li D, Savarese S, Hoi S. BLIP-2: bootstrapping language-image pre-training with frozen image encoders and large language models. In: Proceedings of the 40th international conference on machine learning. PMLR; 2023, p. 19730–42.
[99] Peng Z, Wang W, Dong L, Hao Y, Huang S, Ma S, Wei F. Kosmos-2: grounding multimodal large language models to the world. 2023, arXiv:2306.14824.
[100] Djolonga J, Yung J, Tschannen M, Romijnders R, Beyer L, Kolesnikov A, Puigcerver J, Minderer M, D'Amour A, Moldovan D, Gelly S, Houlsby N, Zhai X, Lucic M. On robustness and transferability of convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 16458–68.
[101] Huang X, Ruan W, Huang W, Jin G, Dong Y, Wu C, Bensalem S, Mu R, Qi Y, Zhao X, Cai K, Zhang Y, Wu S, Xu P, Wu D, Freitas A, Mustafa MA. A survey of safety and trustworthiness of large language models through the lens of verification and validation. Artif Intell Rev 2024;57(7):175.
[102] Li X, Zhang M, Geng Y, Geng H, Long Y, Shen Y, Zhang R, Liu J, Dong H. ManipLLM: embodied multimodal large language model for object-centric robotic manipulation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024, p. 18061–70.
[103] Liang X, Wang Z, Li M, Yan Z. A survey of LLM-augmented knowledge graph construction and application in complex product design. Procedia CIRP 2024;128:870–5.
[104] Xu S, Wei Y, Zheng P, Zhang J, Yu C. LLM enabled generative collaborative design in a mixed reality environment. J Manuf Syst 2024;74:703–15.
[105] Li X, Liu M, Zhang H, Yu C, Xu J, Wu H, Cheang C, Jing Y, Zhang W, Liu H, Li H, Kong T. Vision-language foundation models as effective robot imitators. 2024, arXiv:2311.01378.
[106] Gkournelos C, Konstantinou C, Makris S. An LLM-based approach for enabling seamless human-robot collaboration in assembly. CIRP Ann 2024;73(1):9–12.
[107] Keskin Z, Joosten D, Klasen N, Huber M, Liu C, Drescher B, Schmitt RH. LLM-enhanced human-machine interaction for adaptive decision making in dynamic manufacturing process environments. IEEE Access 2025.
[108] Tian X, Xu J, Wang S, Zhou Y. Construction and application of a multi-modal knowledge graph integrated with large language models in the field of manufacturing processes. In: 2025 international conference on artificial intelligence in information and communication. ICAIIC, IEEE; 2025, p. 0288–92.
[109] Xu J, Chen Z, Ren H, Jiang Z, Wang Y, Gui W. Text-Augmented Contrastive Evaluation Method: Pioneeringly Achieving Quantitative Assessment for Rag-Enhanced Llm of Industrial Fault Diagnosis, Available at SSRN 5146755.
[110] Liu J, Lin F, Li X, Lim KH, Zhao S. Physics-informed LLM-agent for automated modulation design in power electronics systems. 2024, arXiv:2411.14214.
[111] Sun Y, Li X, Liu C, Deng X, Zhang W, Wang J, Zhang Z, Wen T, Song T, Ju D. Development of an intelligent design and simulation aid system for heat treatment processes based on LLM. Mater Des 2024;248:113506.
[112] Fu T, Liu S, Li P. Intelligent smelting process, management system: Efficient and intelligent management strategy by incorporating large language model. Front Eng Manag 2024;11(3):396–412.
[113] Zhong S, Gatti E, Hardwick J, Ribul M, Cho Y, Obrist M. LLM-mediated domain-specific voice agents: The Case of TextileBot. 2024, arXiv:2406.10590.
[114] Zhong S, Gatti E, Cho Y, Obrist M. Feeling Textiles through AI: An exploration into multimodal language models and human perception alignment. In: Proceedings of the 26th international conference on multimodal interaction. ICMI '24, New York, NY, USA: Association for Computing Machinery; 2024, p. 33–7.
[115] Huang W, Wang C, Zhang R, Li Y, Wu J, Fei-Fei L. Voxposer: Composable 3d value maps for robotic manipulation with language models. 2023, arXiv preprint arXiv:2307.05973.
[116] Yang D, Wu A, Zhang T, Zhang L, Liu F, Lian X, Ren Y, Tian J. A multi-agent framework for extensible structured text generation in PLCs. 2024, arXiv:2412.02410.
[117] Liu Z, Zeng R, Wang D, Peng G, Wang J, Liu Q, Liu P, Wang W. Agents4PLC: automating closed-loop PLC code generation and verification in industrial control systems using LLM-based agents. 2024, arXiv:2410.14209.
[118] Pu H, Yang X, Li J, Guo R. AutoRepo: A general framework for multimodal LLM-based automated construction reporting. Expert Syst Appl 2024;255:124601.
[119] Zheng Z, Chen K-Y, Cao X-Y, Lu X-Z, Lin J-R. LLM-FuncMapper: function identification for interpreting complex clauses in building codes via LLM. 2023, arXiv:2308.08728.
[120] Li Z, Xia L, Tang J, Xu Y, Shi L, Xia L, Yin D, Huang C. UrbanGPT: spatio-temporal large language models. In: Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining. KDD '24, New York, NY, USA: Association for Computing Machinery; 2024, p. 5351–62.
[121] Jiang G, Ma Z, Zhang L, Chen J. EPlus-LLM: A large language model-based computing platform for automated building energy modeling. Appl Energy 2024;367:123431.
[122] Stojkovic J, Choukse E, Zhang C, Goiri I, Torrellas J. Towards greener LLMs: bringing energy-efficiency to the forefront of LLM Inference. 2024, arXiv:2403.20306.
[123] Shi F, Zhang Y, Qu C, Fan C, Chu J, Jin L, Liu S. 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, Atlantis Press; 2024, p. 125–33.
[124] Crawford N, Duffy EB, Evazzade I, Foehr T, Robbins G, Saha DK, Varma J, Ziolkowski M. BMW Agents – a framework for task automation through multi-agent collaboration. 2024, arXiv:2406.20041.
[125] Maranto D. LLMSat: a large language model-based goal-oriented agent for autonomous space exploration. 2024, arXiv:2405.01392.
[126] Li Y, Yu X, Koudas N. Data acquisition for improving machine learning models. Proc VLDB Endow 2021;14(10):1832–44, arXiv:2105.14107.
[127] Liu F, Kang Z, Han X. Optimizing RAG techniques for automotive industry PDF chatbots: a case study with locally deployed ollama models. 2024, arXiv:2408.05933.
[128] Klingenberg CO, Borges MAV, Jr. JAVA. Industry 4.0 as a data-driven paradigm: A systematic literature review on technologies. J Manuf Technol Manag 2019;32(3):570–92.
[129] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012;25.
[130] Li G, Yuan C, Kamarthi S, Moghaddam M, Jin X. Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis. J Manuf Syst 2021;60:692–706.
[131] Sun Q, Ge Z. A survey on deep learning for data-driven soft sensors. IEEE Trans Ind Inform 2021;17(9):5853–66.
[132] Izagirre U, Andonegui I, Landa-Torres I, Zurutuza U. A practical and synchronized data acquisition network architecture for industrial robot predictive maintenance in manufacturing assembly lines. Robot Comput-Integr Manuf 2022;74:102287.
[133] Chu X, Ilyas IF, Krishnan S, Wang J. Data cleaning: overview and emerging challenges. In: Proceedings of the 2016 international conference on management of data. SIGMOD '16, New York, NY, USA: Association for Computing Machinery; 2016, p. 2201–6.
[134] Côté P-O, Nikanjam A, Ahmed N, Humeniuk D, Khomh F. Data cleaning and machine learning: a systematic literature review. Autom Softw Eng 2024;31(2):54.
[135] Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O. A survey on missing data in machine learning. J Big Data 2021;8(1):140.
[136] Khan H, Rasheed MT, Liu H, Zhang S. High-order polynomial interpolation with CNN: A robust approach for missing data imputation. Comput Electr Eng 2024;119:109524.
[137] Kwon Y, Zou J. Beta shapley: a unified and noise-reduced data valuation framework for machine learning. 2021, arXiv preprint arXiv:2110.14049.
[138] Adraoui M, Azmi R, Chenal J, Diop EB, Abdem SAE, Serbouti I, Hlal M, Bounabi M. A two-phase approach for leak detection and localization in water distribution systems using wavelet decomposition and machine learning. Comput Ind Eng 2024;197:110534.
[139] Shuai H, Junxia L, Lei W, Wei Z. Research on acoustic fault diagnosis of bearings based on spatial filtering and time-frequency domain filtering. Measurement 2023;221:113533.
[140] Chen H-W, Bandyopadhyay S, Shasha DE, Birnbaum KD. Predicting genome-wide redundancy using machine learning. BMC Evol Biol 2010;10:1–15.
[141] Zhang S, Jafari O, Nagarkar P. A survey on machine learning techniques for auto labeling of video, audio, and text data. 2021, arXiv:2109.03784.
[142] Fredriksson T, Mattos DI, Bosch J, Olsson HH. Data labeling: An empirical investigation into industrial challenges and mitigation strategies. In: International conference on product-focused software process improvement. Springer; 2020, p. 202–16.
[143] Neverova N, Wolf C, Nebout F, Taylor GW. Hand pose estimation through semi-supervised and weakly-supervised learning. Comput Vis Image Underst 2017;164:56–67.

[144] Lin C-W, Chiang C-K, Wang Y-A, Yang Y-L, Li H-T, Lin T-C. ST-LP: Self-training and label propagation for semi-supervised classification. Multimedia Tools Appl 2024;83(41):89335–53.
[145] Ren P, Xiao Y, Chang X, Huang P-Y, Li Z, Gupta BB, Chen X, Wang X. A survey of deep active learning. ACM Comput Surv 2021;54(9):1–40.
[146] Cammarano A, Varriale V, Michelino F, Caputo M. Open and crowd-based platforms: impact on organizational and market performance. Sustainability 2022;14(4):2223.
[147] García S, Luengo J, Herrera F. Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. Knowl-Based Syst 2016;98:1–29.
[148] Singh D, Singh B. Feature wise normalization: An effective way of normalizing data. Pattern Recognit 2022;122:108307.
[149] Singh D, Singh B. Investigating the impact of data normalization on classification performance. Appl Soft Comput 2020;97:105524.
[150] Hancer E, Xue B, Zhang M. A survey on feature selection approaches for clustering. Artif Intell Rev 2020;53(6):4519–45.
[151] Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, et al. On the opportunities and risks of foundation models. 2021, arXiv preprint arXiv:2108.07258.
[152] Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, Yogatama D, Bosma M, Zhou D, Metzler D, et al. Emergent abilities of large language models. 2022, arXiv preprint arXiv:2206.07682.
[153] Qiu X, Sun T, Xu Y, Shao Y, Dai N, Huang X. Pre-trained models for natural language processing: A survey. Sci China Technol Sci 2020;63(10):1872–97.
[154] Li J, Tang T, Zhao WX, Nie J-Y, Wen J-R. Pre-trained language models for text generation: A survey. ACM Comput Surv 2024;56(9):1–39.
[155] Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, et al. A survey on visual transformer. 2020, arXiv preprint arXiv:2012.12556.
[156] Hu W, Liu B, Gomes J, Zitnik M, Liang P, Pande V, Leskovec J. Strategies for pre-training graph neural networks. 2019, arXiv preprint arXiv:1905.12265.
[157] Forsyth DA, Ponce J. Computer vision: A modern approach. Prentice Hall Professional Technical Reference; 2002.
[158] Bondy JA, Murty USR, et al. Graph theory with applications. Vol. 290, Macmillan London; 1976.
[159] Koren Y, Koren Y, et al. Robotics for engineers. Vol. 168, McGraw-Hill New York; 1985.
[160] Liang J, Huang W, Xia F, Xu P, Hausman K, Ichter B, Florence P, Zeng A. Code as policies: Language model programs for embodied control. In: 2023 IEEE international conference on robotics and automation. ICRA, IEEE; 2023, p. 9493–500.
[161] Brohan A, Brown N, Carbajal J, Chebotar Y, Chen X, Choromanski K, Ding T, Driess D, Dubey A, Finn C, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. 2023, arXiv preprint arXiv:2307.15818.
[162] Rae JW, Borgeaud S, Cai T, Millican K, Hoffmann J, Song F, Aslanides J, Henderson S, Ring R, Young S, et al. Scaling language models: Methods, analysis & insights from training gopher. 2021, arXiv preprint arXiv:2112.11446.
[163] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. Language models are few-shot learners. Adv Neural Inf Process Syst 2020;33:1877–901.
[164] Zhang S, Roller S, Goyal N, Artetxe M, Chen M, Chen S, Dewan C, Diab M, Li X, Lin XV, et al. Opt: Open pre-trained transformer language models. 2022, arXiv preprint arXiv:2205.01068.
[165] Kenton JDM-WC, Toutanova LK. Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of naacL-HLT. Vol. 1, Minneapolis, Minnesota; 2019, p. 2.
[166] Liu Y. Roberta: A robustly optimized bert pretraining approach. 2019, arXiv preprint arXiv:1907.11692. 364.
[167] Chen M, Tworek J, Jun H, Yuan Q, Pinto HPDO, Kaplan J, Edwards H, Burda Y, Joseph N, Brockman G, et al. Evaluating large language models trained on code. 2021, arXiv preprint arXiv:2107.03374.
[168] Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. Adv Neural Inf Process Syst 2022;35:22199–213.
[169] Wei J, Wang X, Schuurmans D, Bosma M, Xia F, Chi E, Le QV, Zhou D, et al. Chain-of-thought prompting elicits reasoning in large language models. Adv Neural Inf Process Syst 2022;35:24824–37.
[170] Nair S, Rajeswaran A, Kumar V, Finn C, Gupta A. R3m: A universal visual representation for robot manipulation. 2022, arXiv preprint arXiv:2203.12601.
[171] Majumdar A, Yadav K, Arnaud S, Ma J, Chen C, Silwal S, Jain A, Berges V-P, Wu T, Vakil J, et al. Where are we in the search for an artificial visual cortex for embodied intelligence? Adv Neural Inf Process Syst 2023;36:655–77.
[172] Yang J, Gao M, Li Z, Gao S, Wang F, Zheng F. Track anything: Segment anything meets videos. 2023, arXiv preprint arXiv:2304.11968.
[173] Zhang C, Han D, Qiao Y, Kim JU, Bae S-H, Lee S, Hong CS. Faster segment anything: Towards lightweight sam for mobile applications. 2023, arXiv preprint arXiv:2306.14289.
[174] Zhang D, Yu Y, Dong J, Li C, Su D, Chu C, Yu D. Mm-llms: Recent advances in multimodal large language models. 2024, arXiv preprint arXiv:2401.13601.
[175] Wang J, Tian Y, Wang Y, Yang J, Wang X, Wang S, Kwan O. A framework and operational procedures for metaverses-based industrial foundation models. IEEE Trans Syst Man Cybern: Syst 2022;53(4):2037–46.
[176] Yin S, Fu C, Zhao S, Li K, Sun X, Xu T, Chen E. A survey on multimodal large language models. 2023, arXiv preprint arXiv:2306.13549.
[177] Caffagni D, Cocchi F, Barsellotti L, Moratelli N, Sarto S, Baraldi L, Cornia M, Cucchiara R. The (r) evolution of multimodal large language models: A survey. 2024, arXiv preprint arXiv:2402.12451.
[178] Koch J, Jevremovic D, Moenck K, Schüppstuhl T. A digital assistance system leveraging vision foundation models & 3D localization for reproducible defect segmentation in visual inspection. Procedia CIRP 2024;130:387–97.
[179] Driess D, Xia F, Sajjadi MS, Lynch C, Chowdhery A, Wahid A, Tompson J, Vuong Q, Yu T, Huang W, et al. Palm-e: An embodied multimodal language model. 2023.
[180] Liu S, Wang L. Vision intelligence-conditioned reinforcement learning for precision assembly. CIRP Annals 2025.
[181] Xu J, Wu H, Wang J, Long M. Anomaly transformer: Time series anomaly detection with association discrepancy. 2021, arXiv preprint arXiv:2110.02642.
[182] Gao J, Shen T, Wang Z, Chen W, Yin K, Li D, Litany O, Gojcic Z, Fidler S. Get3d: A generative model of high quality 3d textured shapes learned from images. Adv Neural Inf Process Syst 2022;35:31841–54.
[183] Zhang J, Liu S, Gao RX, Wang L. Neural rendering-enabled 3d modeling for rapid digitization of in-service products. CIRP annals 2023;72(1):93–6.
[184] Sun Y, Zhang Q, Bao J, Lu Y, Liu S. Empowering digital twins with large language models for global temporal feature learning. J Manuf Syst 2024;74:83–99.
[185] Liu S, Guo D, Liu Z, Wang T, Qin Q, Wang XV, Wang L. A digital twin-enabled approach to reliable human–robot collaborative assembly. In: Human-Centric Smart Manufacturing Towards Industry 5.0. Springer; 2025, p. 281–304.
[186] Wang G, Zhang C, Liu S, Zhao Y, Zhang Y, Wang L. Multi-robot collaborative manufacturing driven by digital twins: advancements, challenges, and future directions. Journal of Manufacturing Systems 2025;82:333–61.
[187] Zhang J, Wang Y, Molino P, Li L, Ebert DS. Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models. IEEE Trans Vis Comput Graphics 2018;25(1):364–73.
[188] Luo J, Xu C, Wu J, Levine S. Precise and dexterous robotic manipulation via human-in-the-loop reinforcement learning. 2024, arXiv preprint arXiv:2410.21845.
[189] Gao F, Xia L, Zhang J, Liu S, Wang L, Gao RX. Integrating large language model for natural language-based instruction toward robust human-robot collaboration. Procedia CIRP 2024;130:313–8.
[190] Liu S, Zhang J, Gao RX, Wang XV, Wang L. Vision-language model-driven scene understanding and robotic object manipulation. In: 2024 IEEE 20th international conference on automation science and engineering. CASE, IEEE; 2024, p. 21–6.
[191] OpenAI Community. OpenAI ChatGPT robot figure 01. 2024, URL https://community.openai.com/t/openai-chatgpt-robot-figure-01/681733. [Accessed 27 March 2025].
[192] Liu S, Zhang J, Wang L, Gao RX. Vision AI-based human-robot collaborative assembly driven by autonomous robots. CIRP Ann 2024;73(1):13–6.
[193] Li C, Gan Z, Yang Z, Yang J, Li L, Wang L, Gao J, et al. Multimodal foundation models: From specialists to general-purpose assistants. Found Trends® Comput Graph Vis 2024;16(1–2):1–214.
[194] Van Engelen JE, Hoos HH. A survey on semi-supervised learning. Mach Learn 2020;109(2):373–440.
[195] Learning S-S. Semi-supervised learning. CSZ2006. Html 2006;5:2.
[196] Zhou Z-H. A brief introduction to weakly supervised learning. Natl Sci Rev 2018;5(1):44–53.
[197] Zhai X, Oliver A, Kolesnikov A, Beyer L. S4l: Self-supervised semi-supervised learning. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019, p. 1476–85.
[198] Zhao B, Fan K, Yang K, Wang Z, Li H, Yang Y. Anonymous and privacy-preserving federated learning with industrial big data. IEEE Trans Ind Inform 2021;17(9):6314–23.
[199] Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: A survey. Int J Robot Res 2013;32(11):1238–74.
[200] Awais M, Naseer M, Khan S, Anwer RM, Cholakkal H, Shah M, Yang M-H, Khan FS. Foundation models defining a new era in vision: a survey and outlook. IEEE Trans Pattern Anal Mach Intell 2025.
[201] Gu J, Han Z, Chen S, Beirami A, He B, Zhang G, Liao R, Qin Y, Tresp V, Torr P. A systematic survey of prompt engineering on vision-language foundation models. 2023, arXiv preprint arXiv:2307.12980.
[202] Ye Q, Axmed M, Pryzant R, Khani F. Prompt engineering a prompt engineer. 2023, arXiv preprint arXiv:2311.05661.
[203] Chen B, Zhang Z, Langrené N, Zhu S. Unleashing the potential of prompt engineering in large language models: a comprehensive review. 2023, arXiv preprint arXiv:2310.14735.
[204] Vatsal S, Dubey H. A survey of prompt engineering methods in large language models for different nlp tasks. 2024, arXiv preprint arXiv:2407.12994.
[205] White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J, Schmidt DC. A prompt pattern catalog to enhance prompt engineering with chatgpt. 2023, arXiv preprint arXiv:2302.11382.
[206] DAIRAI. Prompt engineering guide. 2024, URL https://www.promptingguide.ai/. [Accessed 14 January 2025].

[207] Rawte V, Sheth A, Das A. A survey of hallucination in large foundation models. 2023, arXiv:2309.05922.
[208] Zhao P, Zhang H, Yu Q, Wang Z, Geng Y, Fu F, Yang L, Zhang W, Jiang J, Cui B. Retrieval-augmented generation for AI-generated content: a survey. 2024, arXiv:2402.19473.
[209] Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, Dai Y, Sun J, Wang M, Wang H. Retrieval-augmented generation for large language models: a survey. 2024, arXiv:2312.10997.
[210] Jiang Z, Xu FF, Gao L, Sun Z, Liu Q, Dwivedi-Yu J, Yang Y, Callan J, Neubig G. Active retrieval augmented generation. 2023, arXiv:2305.06983.
[211] Guo L, Yan F, Li T, Yang T, Lu Y. An automatic method for constructing machining process knowledge base from knowledge graph. Robot Comput-Integr Manuf 2022;73:102222.
[212] Cao Q, Zanni-Merk C, Samet A, Reich C, de Beuvron FdB, Beckmann A, Giannetti C. KSPMI: a knowledge-based system for predictive maintenance in industry 4.0. Robot Comput-Integr Manuf 2022;74:102281.
[213] Singh PN, Talasila S, Banakar SV. Analyzing embedding models for embedding vectors in vector databases. In: 2023 IEEE international conference on ICT in business industry & government. ICTBIG, 2023, p. 1–7.
[214] Askari A, Abolghasemi A, Pasi G, Kraaij W, Verberne S. Injecting the BM25 Score as text improves BERT-based re-rankers. In: Kamps J, Goeuriot L, Crestani F, Maistro M, Joho H, Davis B, Gurrin C, Kruschwitz U, Caputo A, editors. Advances in information retrieval. Cham: Springer Nature Switzerland; 2023, p. 66–83.
[215] Zhao WX, Liu J, Ren R, Wen J-R. Dense text retrieval based on pretrained language models: a survey. ACM Trans Inf Syst 2024;42(4):89:1–60.
[216] Chan C-M, Xu C, Yuan R, Luo H, Xue W, Guo Y, Fu J. RQ-RAG: learning to refine queries for retrieval augmented generation. 2024, arXiv:2404.00610.
[217] Madhav D, Nijai S, Patel U, Champanerkar K. Question generation from PDF using LangChain. In: 2024 11th international conference on computing for sustainable global development. INDIACom, 2024, p. 218–22.
[218] Abyaneh MB, Vahdani B, Nadjafi BA, Amiri M. Integrated scheduling of multiple heterogeneous equipment and maintenance operations during discharging process in a container terminal under uncertainty. Comput Ind Eng 2024;193:110300.
[219] Zhao S, Kang K, Xu C, Guo X, Zhong RY. Digital twin enabled construction site monitoring (CSM) method with edge-cloud collaboration. In: 2024 IEEE 20th international conference on automation science and engineering. CASE, 2024, p. 3017–22.
[220] Wang J, Tian Y, Wang Y, Yang J, Wang X, Wang S, Kwan O. A framework and operational procedures for metaverses-based industrial foundation models. IEEE Trans Syst Man Cybern: Syst 2023;53(4):2037–46.
[221] Latsou C, Farsi M, Erkoyuncu JA. Digital twin-enabled automated anomaly detection and bottleneck identification in complex manufacturing systems using a multi-agent approach. J Manuf Syst 2023;67:242–64.
[222] Yang S, Nachum O, Du Y, Wei J, Abbeel P, Schuurmans D. Foundation models for decision making: problems, methods, and opportunities. 2023, arXiv:2303.04129.
[223] Shen W, Wang L, Hao Q. Agent-based distributed manufacturing process planning and scheduling: A state-of-the-art survey. IEEE Trans Syst Man Cybern Part C (Appl Rev.) 2006;36(4):563–77.
[224] Sado F, Loo CK, Liew WS, Kerzel M, Wermter S. Explainable goal-driven agents and robots - a comprehensive review. ACM Comput Surv 2023;55(10):211:1–41.
[225] Xu R, Li J, Dong X, Yu H, Ma J. Bridging the domain gap for multi-agent perception. In: 2023 IEEE international conference on robotics and automation. ICRA, 2023, p. 6035–42.
[226] Huang X, Liu W, Chen X, Wang X, Wang H, Lian D, Wang Y, Tang R, Chen E. Understanding the planning of LLM agents: A survey. 2024, arXiv:2402.02716.
[227] Zhao Y, Wang J, Xiang L, Zhang X, Guo Z, Turkay C, Zhang Y, Chen S. LightVA: lightweight visual analytics with LLM agent-based task planning and execution. IEEE Trans Vis Comput Graphics 2024;1–13.
[228] Feng G, Zhang B, Gu Y, Ye H, He D, Wang L. Towards revealing the mystery behind chain of thought: a theoretical perspective. Adv Neural Inf Process Syst 2023;36:70757–98.
[229] Sun R, Wang Y, Mai H, Zhang T, Wu F. Alignment before aggregation: trajectory memory retrieval network for video object segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 1218–28.
[230] Zhao S, Wang J, Zhang J, Bao J, Zhong R. Edge-cloud collaborative fabric defect detection based on industrial internet architecture. In: 2020 IEEE 18th international conference on industrial informatics. INDIN, Vol. 1, 2020, p. 483–7.
[231] Lim J, Vogel-Heuser B, Kovalenko I. Large language model-enabled multi-agent manufacturing systems. 2024, arXiv:2406.01893.
[232] Golpayegani F, Chen N, Afraz N, Gyamfi E, Malekjafarian A, Schäfer D, Krupitzer C. Adaptation in edge computing: a review on design principles and research challenges. ACM Trans Auton Adapt Syst 2024;19(3):19:1–43.
[233] Shen Y, Song K, Tan X, Li D, Lu W, Zhuang Y. HuggingGPT: solving AI tasks with ChatGPT and its friends in hugging face. Adv Neural Inf Process Syst 2023;36:38154–80.
[234] Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E, Le QV, Zhou D. Chain-of-thought prompting elicits reasoning in large language models. Adv Neural Inf Process Syst 2022;35:24824–37.
[235] Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K, Cao Y. ReAct: synergizing reasoning and acting in language models. 2023, arXiv:2210.03629.
[236] Wang X, Wei J, Schuurmans D, Le Q, Chi E, Narang S, Chowdhery A, Zhou D. Self-consistency improves chain of thought reasoning in language models. 2023, arXiv:2203.11171.
[237] Yao S, Yu D, Zhao J, Shafran I, Griffiths T, Cao Y, Narasimhan K. Tree of thoughts: deliberate problem solving with large language models. Adv Neural Inf Process Syst 2023;36:11809–22.
[238] Dagan G, Keller F, Lascarides A. Dynamic Planning with a LLM. 2023, arXiv:2308.06391.
[239] Chen L, Lu K, Rajeswaran A, Lee K, Grover A, Laskin M, Abbeel P, Srinivas A, Mordatch I. Decision transformer: reinforcement learning via sequence modeling. In: Advances in neural information processing systems. Vol. 34, Curran Associates, Inc.; 2021, p. 15084–97.
[240] Zhang D, Chen L, Zhang S, Xu H, Zhao Z, Yu K. Large language models are semi-parametric reinforcement learning agents. Adv Neural Inf Process Syst 2023;36:78227–39.
[241] Zhong W, Guo L, Gao Q, Ye H, Wang Y. MemoryBank: enhancing large language models with long-term memory. Proc AAAI Conf Artif Intell 2024;38(17):19724–31.
[242] Wang Y, Jiang Z, Chen Z, Yang F, Zhou Y, Cho E, Fan X, Huang X, Lu Y, Yang Y. RecMind: large language model powered agent for recommendation. 2024, arXiv:2308.14296.
[243] Park JS, O'Brien J, Cai CJ, Morris MR, Liang P, Bernstein MS. Generative agents: interactive simulacra of human behavior. In: Proceedings of the 36th annual ACM symposium on user interface software and technology. UIST '23, New York, NY, USA: Association for Computing Machinery; 2023, p. 1–22.
[244] Cho J, Yoon J, Ahn S. Spatially-aware transformer for embodied agents. 2024, arXiv:2402.15160.
[245] Wang L, Ma C, Feng X, Zhang Z, Yang H, Zhang J, Chen Z, Tang J, Chen X, Lin Y, Zhao WX, Wei Z, Wen J. A survey on large language model based autonomous agents. Front Comput Sci 2024;18(6):186345.
[246] Zhang Y, Ma Z, Ma Y, Han Z, Wu Y, Tresp V. WebPilot: a versatile and autonomous multi-agent system for web task execution with strategic exploration. 2024, arXiv:2408.15978.
[247] Harish A, Prakash G, Nair RR, Iyer VB, M AK. Refining LLMs with reinforcement learning for human-like text generation. In: 2024 IEEE international conference on electronics, computing and communication technologies. CONECCT, 2024, p. 1–6.
[248] Zhao S, Zhong RY, Xu C, Wang J, Zhang J. A dynamic inference network (DI-Net) for online fabric defect detection in smart manufacturing. J Intell Manuf 2024.
[249] Zheng P, Li S, Fan J, Li C, Wang L. A collaborative intelligence-based approach for handling human-robot collaboration uncertainties. CIRP Ann 2023;72(1):1–4.
[250] Fan J, Zheng P, Li S. Vision-based holistic scene understanding towards proactive human–robot collaboration. Robot Comput-Integr Manuf 2022;75:102304.
[251] Wang B, Zheng P, Yin Y, Shih A, Wang L. Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective. J Manuf Syst 2022;63:471–90.
[252] Wang L, Gao RX, Krüger J, Váncza J. Human-centric assembly in smart factories. CIRP Annals 2025.
[253] Wang L. A futuristic perspective on human-centric assembly. J Manuf Syst 2022;62:199–201.
[254] Fan H, Liu X, Fuh JYH, Lu WF, Li B. Embodied intelligence in manufacturing: Leveraging large language models for autonomous industrial robotics. J Intell Manuf 2024.
[255] Liu S, Wang L, Vincent Wang X. Multimodal data-driven robot control for human–robot collaborative assembly. Journal of Manufacturing Science and Engineering 2022;144(5):051012.
[256] Wang T, Fan J, Zheng P. An LLM-based vision and language cobot navigation approach for human-centric smart manufacturing. J Manuf Syst 2024;75:299–305.
[257] Ni M, Wang T, Leng J, Chen C, Cheng L. A large language model-based manufacturing process planning approach under industry 5.0. Int J Prod Res 2025;1–20.
[258] Xu Q, Zhou G, Zhang C, Chang F, Cao Y, Zhao D. Generative AI and DT integrated intelligent process planning: a conceptual framework. Int J Adv Manuf Technol 2024;133(5):2461–85.
[259] Yi S, Liu S, Yang Y, Yan S, Guo D, Wang XV, Wang L. Safety-aware human-centric collaborative assembly. Adv Eng Inform 2024;60:102371.
[260] Liu S, Wang L, Wang XV. Sensorless force estimation for industrial robots using disturbance observer and neural learning of friction approximation. Robotics and Computer-Integrated Manufacturing 2021;71:102168.
[261] Liu S, Wang L, Wang XV. Sensorless haptic control for human-robot collaborative assembly. CIRP Journal of Manufacturing Science and Technology 2021;32:132–44.

[262] Ji Y, Zhang Z, Tang D, Zheng Y, Liu C, Zhao Z, Li X. Foundation models assist in human–robot collaboration assembly. Sci Rep 2024;14(1):24828.
[263] Joglekar O, Kozlovsky S, Lancewicki T, Tchuiev V, Feldman Z, Di Castro D. Towards natural language-driven industrial assembly using foundation models. In: ICLR 2024 workshop on large language model (LLM) agents. 2024.
[264] Gui Y, Tang D, Zhu H, Zhang Y, Zhang Z. Dynamic scheduling for flexible job shop using a deep reinforcement learning approach. Comput Ind Eng 2023;180:109255.
[265] Qin Z, Johnson D, Lu Y. Dynamic production scheduling towards self-organizing mass personalization: A multi-agent dueling deep reinforcement learning approach. J Manuf Syst 2023;68:242–57.
[266] Yu J, Liu J. Multiple granularities generative adversarial network for recognition of wafer map defects. IEEE Trans Ind Inform 2022;18(3):1674–83.
[267] Azamfirei V, Psarommatis F, Lagrosen Y. Application of automation for in-line quality inspection, a zero-defect manufacturing approach. J Manuf Syst 2023;67:1–22.
[268] Wang H, Li C, Li Y-F. Large-scale visual language model boosted by contrast domain adaptation for intelligent industrial visual monitoring. IEEE Trans Ind Inform 2024;20(12):14114–23.
[269] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, arXiv preprint arXiv:1409.1556.
[270] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8.
[271] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 4700–8.
[272] Zhu Y, Zhu M, Liu N, Xu Z, Peng Y. Llava-phi: Efficient multi-modal assistant with small language model. In: Proceedings of the 1st international workshop on efficient multimedia computing under limited. 2024, p. 18–22.
[273] Xu Q, Qiu F, Zhou G, Zhang C, Ding K, Chang F, Lu F, Yu Y, Ma D, Liu J. A large language model-enabled machining process knowledge graph construction method for intelligent process planning. Adv Eng Inform 2025;65:103244.
[274] Zhang C, Xu Q, Yu Y, Zhou G, Zeng K, Chang F, Ding K. A survey on potentials, pathways and challenges of large language models in new-generation intelligent manufacturing. Robot Comput-Integr Manuf 2025;92:102883.