ENGINEERING Management 2026, 13(1): 42–64
https://doi.org/10.1007/s42524-026-5273-5

REVIEW ARTICLE

Xinyu PAN, Di WANG, Fugee TSUNG

Empowering intelligent quality control with large models: A comprehensive survey of methods, challenges, and perspectives

© Higher Education Press 2026

Abstract Quality control (QC) serves as a cornerstone of modern manufacturing, exerting a decisive influence on production efficiency, product reliability and customer satisfaction. However, traditional QC systems, which largely rely on rule-based frameworks and narrowly defined statistical methods, face increasing limitations in handling the scale, diversity and complexity of contemporary industrial data. This limitation provides a strong motivation to explore the potential of large models (LMs) for advancing QC. Distinguished by their powerful capabilities in knowledge integration, contextual understanding and adaptive reasoning, LMs offer transformative opportunities to modernize QC. This review begins by analyzing why LMs are particularly well positioned to enhance QC, focusing on three crucial dimensions: input alignment, which enables seamless integration of heterogeneous data sources; task adaptability, which supports associative learning across multiple QC tasks and allows knowledge transfer; and augmented intelligence, which supports human experts in complex decision-making. Recent advances in industrial applications are summarized, with particular attention to methodological innovations, deployment practices and integration pathways into manufacturing workflows. To systematically structure the current landscape, the key challenges are categorized into three interrelated dimensions, i.e., data, model and evaluation, which correspond to the core requirements for model training, practical implementation and sustainable adaptability in real-world scenarios. Building on this foundation, the review further outlines future research directions, highlighting secure data collaboration, system-level integration and continual learning under dynamic environments as critical priorities for the next stage of development. Collectively, these insights underscore the promise of LMs in reshaping QC into an intelligent, resilient and future-ready paradigm.

Keywords quality control, large models, foundation models, industrial artificial intelligence survey, manufacturing intelligence

1 Introduction

In the era of intelligent manufacturing, quality control (QC) has transcended its traditional role as a post-production checkpoint and become a strategic component of the entire manufacturing lifecycle. It is central to maintaining the stability of production processes, ensuring the consistency of product quality and supporting the overall resilience of manufacturing systems. Over the past century, QC has undergone several significant evolutionary stages. It originated as empirical quality inspection in the early 20th century, advanced to statistical quality control (SQC) with the introduction of Shewhart control charts and later, evolved into SPC and total quality management (TQM), emphasizing organization-wide quality awareness (Ebadi et al., 2021; Woodall and Montgomery, 2014). With the advent of digitalization and cyber-physical systems under the Industry 4.0 paradigm, quality QC has entered the era of "Quality 4.0," characterized by

Received Sep. 12, 2025; revised Dec. 7, 2025; accepted Jan. 4, 2026

Xinyu PAN, Di WANG (✉)
Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai 200240, China
E-mail: d.wang@sjtu.edu.cn

Fugee TSUNG
Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Hong Kong SAR 999077, China

This work was supported by the National Natural Science Foundation of China (Grant Nos. 72471145, 72371217, and 72101148), the Youth Talent Support Program of the China Association for Science and Technology (No. YESS20240068), Guangzhou Industrial Informatics and Intelligence Key Laboratory, China (No. 2024A03J0628), and Nansha Key Area Science and Technology Project, China (Nos. 2023ZD003 and 2021JC02X191).

increasingly complex, interconnected, and data-intensive manufacturing environments where the artificial intelligence (AI), industrial Internet of things (IIoTs), and advanced analytics are leveraged to build intelligent and adaptive quality systems. (Escobar et al., 2025; Klingenberg et al., 2019; Lee et al., 2019). This new stage of QC departs from the traditionally reactive paradigm and instead emphasizes proactive and predictive strategies, shifting the focus from post-failure remediation to quality prevention and early-stage prediction (Megahed et al., 2024; Gomaa, 2025). In this context, modern QC should be redefined as an intelligent and data-driven process spanning the entire product lifecycle: quality-oriented product and process design, which involves learning from historical failures and optimize design parameters (Yu et al., 2021; Deng et al., 2023); in-process monitoring, which utilizes sensor fusion to ensure production consistency (Woodall and Montgomery, 2014; Yin et al., 2014); intelligent fault diagnosis, which identifies the root causes of anomalies and defects (Kouchakzadeh and ElMaraghy, 2024; Lei et al., 2020); and predictive health maintenance, which anticipates potential degradations or failures based on historical and real-time data and reduce unexpected downtime through continuous prognostics (Hu et al., 2022; Gomaa, 2025).

Along this trajectory, several method lines have recently advanced these tasks from different angles. Bayesian and advanced statistical approaches extend SPC to high-dimensional and dynamic production data by embedding prior process knowledge in hierarchical or nonparametric models, yielding interpretable posterior risk measures for decision-making (Yang and Zhang, 2024; Qiu and Xie, 2022); they also enable multichannel profile/functional monitoring (Capezza et al., 2024) and adaptive EWMA-type charts that track complex trajectories (Capezza et al., 2025), as well as online Bayesian change-point detection under partial observability and strong cross-sensor dependence (Guo et al., 2023). In parallel, digital-twin (DT) methods couple physics-based simulation with streaming data to maintain an evolving virtual counterpart of products and equipment, supporting real-time state estimation, process-parameter what-if analysis, and multi-scale quality characterization in machining (Gaikwad et al., 2020; Psarommatis and May, 2023; Liu et al., 2023a); DT-enabled predictive maintenance further integrates degradation models and virtual tests to mitigate scarce failure data and to close the loop from diagnosis to control (van Dinter et al., 2022). Meanwhile, data-driven machine learning (ML) and deep learning (DL) methods, offer automated feature learning for inspection and prediction (Li et al., 2024a; Lv et al., 2024), such as multistage multi-task networks that jointly model variation propagation and multiple quality indices (Wang et al., 2023a), and multisensor-fusion models for time-series quality forecasting and anomaly detection (Tercan and Meisen, 2022; Wu et al., 2022). However, these approaches are often limited by their reliance on single-modality data and narrowly defined tasks, making them incapable of handling the complexities of modern manufacturing environments. They struggle with scaling to large and integrated systems, require substantial amounts of labeled data for effective performance, and lack robustness when dealing with high-dimensional, heterogeneous, or previously unseen conditions. These limitations highlight an urgent need for more scalable, flexible, and generalizable solutions capable of satisfying the dynamic and multifaceted demands of intelligent manufacturing.

While industrial QC is evolving toward data-driven practices, a parallel revolution is taking place within the broader field of AI (Bommasani et al., 2022). Large models (LMs), including large language models (LLMs), large multimodal models (LMMs), large vision-language models (LVLMs) and so on, represent not just incremental improvements over traditional ML approaches, but rather a fundamental shift in how AI learns and generalizes. LMs stand out primarily due to their massive scale in both data and parameters. These models are pretrained on vast and diverse data sets, ranging from text corpora and image collections to multimodal sources, allowing them to develop a rich and deep contextual understanding. This scale enables LMs to excel in tasks like few-shot and zero-shot learning, where they can make predictions or generate outputs with minimal task-specific examples. Central to the power of these models is the Transformer architecture (Vaswani et al., 2017), which uses attention mechanisms to capture long-range dependencies, understand nuanced semantics, and model cross modality relationships effectively. By focusing on relevant parts of the input data, transformers enable LMs to learn intricate patterns and relationships, leading to more accurate and adaptable AI systems. Consequently, LMs have demonstrated remarkable performance across various fields. In natural language processing (NLP), they enable sophisticated language understanding and generation, owing to knowledge captured from massive text corpora (Devlin et al., 2019; Brown et al., 2020). In computer vision (CV), they surpass traditional convolutional neural networks (CNNs) in tasks like image classification, segmentation and visual reasoning (Kawaharazuka et al., 2023). In the healthcare domain, they are becoming a cornerstone of precision medicine by revolutionizing diagnostics, i.e., analyzing pathology images, genomic data and electronic health records to detect anomalies and guide personalized treatments (Mahesh et al., 2024). In summary, LMs have shown transformative capabilities, delivering state-of-the-art performance across diverse domains. Their adoption enables manufacturing to evolve from rule-based inspection to AI-driven systems that capture complex patterns, detect anomalies with high precision and provide

real-time prescriptive insights (Li et al., 2024b; Zhang et al., 2026). These advances lay the foundation for a new paradigm in industrial AI.

As illustrated in Fig. 1, the advancements clearly indicate that LMs hold great promise for transforming QC within today's data-intensive manufacturing landscape. Conversely, the unique and complex requirements posed by industrial scenarios can, in turn, drive the advancement of LMs by inspiring new capabilities, adaptations and research directions tailored to real-world applications. However, their successful integration necessitates a deeper understanding of both their capabilities and limitations. This review makes three primary contributions. First, we explore the fundamental reasons why LMs have the potential to empower QC tasks. Second, we conduct a comprehensive survey of current research efforts that have applied LMs to real-world QC scenarios, structured around three core functional dimensions: perception, reasoning and interaction, mirroring the end-to-end lifecycle of quality-related data. Third, we highlight the key challenges that still hinder the effective deployment of LMs in industrial QC and outline future perspectives that may help overcome these barriers. Ultimately, our objective is to bridge the current disconnect between the theoretical capabilities of LMs and their tangible industrial applications in quality enhancement.

2 Motivation for large models in quality control

The adoption of LMs in QC is not merely a response to technical limitations, but a natural outcome of the evolving landscape of modern manufacturing. As QC processes increasingly generate multimodal data and as inspection tasks grow more interdependent, there is a rising demand for models with stronger generalization, adaptability and intelligence. This section highlights three key drivers behind the integration of LMs into QC (see Fig. 2): Input alignment reflects the ability of LMs to effectively handle temporal, multimodal and low-label manufacturing data; Task adaptability emphasizes their capacity to generalize across designing, monitoring, diagnosis and predictive health prognostics tasks through unified representation and joint modeling; Augmented intelligence captures how LMs integrate domain knowledge with reasoning and generation capabilities to enhance interpretability and decision support in QC processes.

2.1 Input alignment

The rapid digitalization of manufacturing has led to the accumulation of massive, diverse and high-dimensional data, which span time-series signals, text and sensor

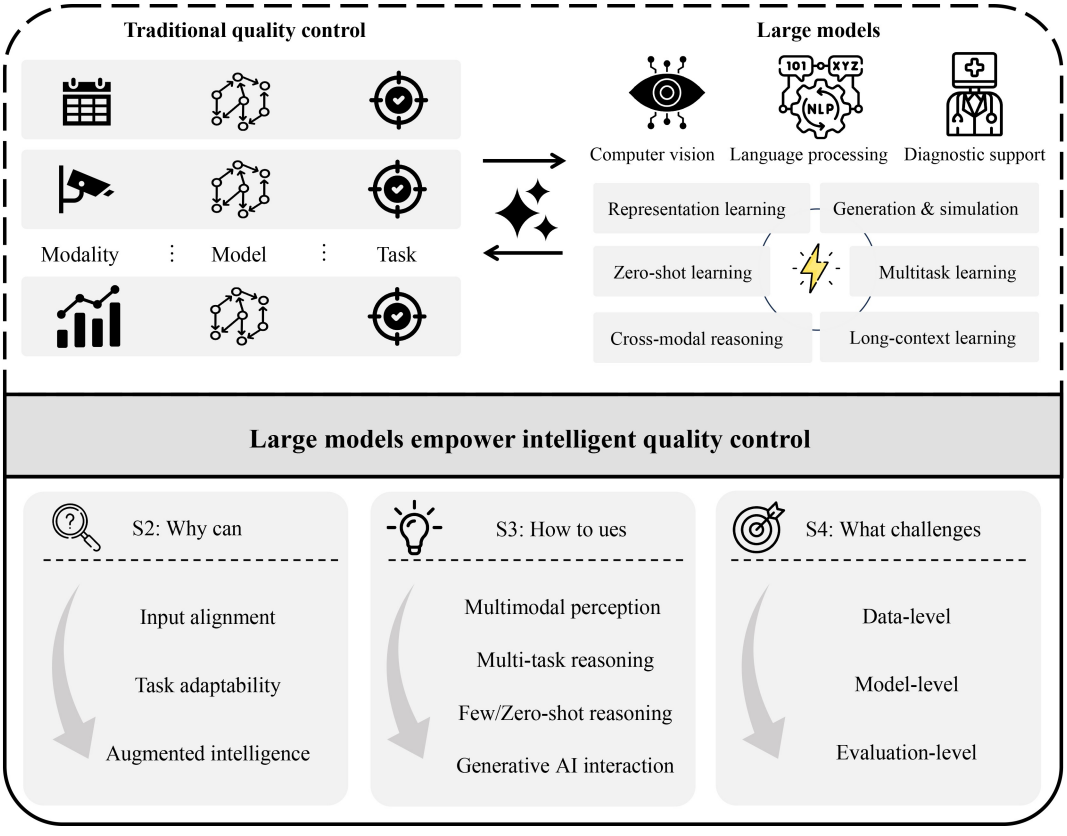

Fig. 1 Igniting innovation at the crossroads of quality control and large models.

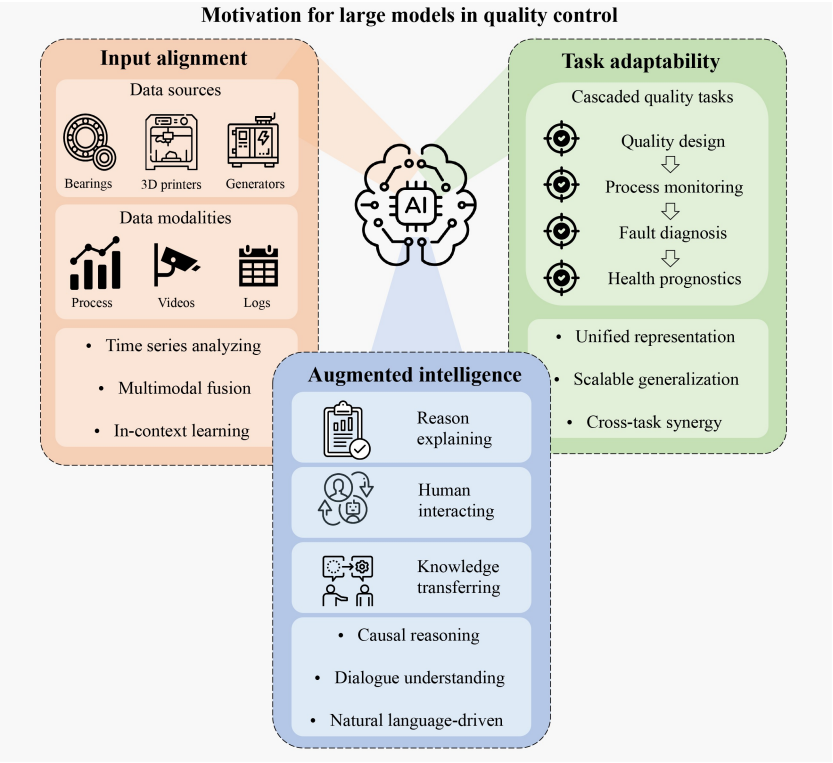

Fig. 2 Motivation for leveraging large models in quality control.

streams. The data richness provides a natural foundation for LMs to demonstrate their strengths in unified representation, enabling more robust and scalable QC.

2.1.1 Temporal reasoning

Manufacturing processes generate vast amounts of time-series data, including sensor readings and machine parameters, where quality outcomes are influenced by intricate temporal patterns and long-term dependencies. Time series analysis enables manufacturers to monitor quality data over time, identifying patterns, trends and anomalies to maintain consistent production standards (Fatima and Rahimi, 2024). However, traditional models like ARIMA (Ho and Xie, 1998) often fall short in capturing the nonlinear effects inherent in such data. While ML methods offer improved capacity to capture nonlinear patterns, they still rely heavily on handcrafted features and struggle with modeling long-range dependencies inherent in complex temporal processes. These limitations become more pronounced as data scales and task complexity grow.

In contrast, LMs, particularly LLMs, have demonstrated a remarkable capacity to capture long-range dependencies in sequential data. Originating in NLP, they convert raw text into discrete tokens using subword methods such as BPE, WordPiece, and SentencePiece (Sennrich et al., 2016; Schuster and Nakajima, 2012; Kudo and Richardson, 2018). Once tokenized and embedded, these sequences are processed by the Transformer architecture, in which attention-based computations allow each token to access information from any other position in the sequence. This design overcomes the locality and memory constraints of traditional recurrent or convolutional models, enabling the model to learn hierarchical structure, contextual relationships, and long-range semantic dependencies that are difficult for earlier DL architectures to capture. Motivated by this success in language, recent research has sought to represent real-valued time series in an analogous tokenized form, allowing them to benefit from the same modeling capabilities. Continuous signals can be discretized into “time-series tokens” using several strategies, including vector-quantized codebooks that assign temporal segments to learned discrete indices (Talukder et al., 2024), patch-based segmentation that embeds fixed-length windows across multiple temporal scales (Peršak et al., 2024), and wavelet-based tokenization that quantizes time–frequency coefficients into compact vocabularies (Masserano et al., 2024).

Building on these developments, recent studies have shown that LMs exhibit strong temporal-reasoning abilities across a wide range of domains. By treating sequential data as tokenized blocks, these models learn hierarchical and contextual representations directly from raw sequences, enabling them to capture both fine-grained local dynamics and long-horizon temporal structure that traditional DL methods struggle to model (Su et al., 2024). Moreover, the attention-based computations

within Transformers further allow the model to selectively emphasize relevant temporal relationships, yielding robust performance in long-range forecasting, anomaly detection, and complex multivariate sequence modeling (Wen et al., 2022; Ahmed et al., 2023; Kim et al., 2024). Beyond architectural expressiveness, LMs trained on extensive time-series corpora achieve highly competitive results. For instance, Time-LLM re-encode numerical sequences into token sequences aligned with textual prototypes, allowing even frozen LLMs to perform competitive forecasting (Jin et al., 2024b). Evidence from recent time-series LMs further indicates that their high capacity, expressive token-based representations, and strong cross-domain generalization often allow them to match or surpass specialized architectures on a wide spectrum of predictive and preventive tasks (Rasul et al., 2023). Collectively, these strengths make LMs well positioned for transfer to industrial QC scenarios (Kottapalli et al., 2025). They can support not only essential time-series analysis tasks such as process monitoring and fault diagnosis but also more proactive, predictive applications including quality-parameter design, early-warning assessment, and predictive health maintenance, indicating the potential of LM-based temporal reasoning to serve as a unified modeling paradigm for next-generation QC systems.

2.1.2 Multimodal fusion

QC in modern factories involves multimodal data, including high-resolution images of products, time-series from sensors (vibration, temperature), textual logs and even audio signals from machines (McKinney et al., 2025). Each modality offers unique and complementary insights into product quality, failure modes, operational anomalies and equipment health conditions. However, the heterogeneity and complexity of these data pose significant challenges to conventional ML models, which often struggle with modality alignment, feature fusion and robust generalization across varying production conditions (Baltrušaitis et al., 2019).

Recent advances in LMMs have demonstrated transformative capabilities in learning unified representations from diverse modalities. Unlike traditional multimodal pipelines that rely on early fusion of raw features, mid-level fusion in shared latent spaces, or late fusion at the decision level (Atrey et al., 2010), LMMs are typically built upon Transformer architectures equipped with cross-modal attention mechanisms. Models such as ViLBERT (Lu et al., 2019) and LXMERT (Tan and Bansal, 2019) introduce co-attentional or cross-modality Transformer layers that enable fine-grained interactions between visual and textual tokens, thereby capturing nuanced inter-modal correspondences and improving semantic alignment. Multimodal Transformer like MulT further shows that cross-modal attention can explicitly model temporal dependencies and correlations in unaligned multimodal sequences (Tsai et al., 2019). The success of Flamingo (Alayrac et al., 2022), which integrates visual features into an LLM backbone via gated cross-attention layers, further illustrates the effectiveness of such architectures in open-world scenarios. These mechanisms enable the model to selectively attend to semantically relevant cues across modalities, establish deeper and more stable alignment of heterogeneous signals, and ultimately map them into a shared embedding space that supports complex multimodal reasoning.

Building on these architectural advances, pioneering works such as CLIP (Radford et al., 2021), ALIGN (Jia et al., 2021), and LiT (Zhai et al., 2022) have shown that training on hundreds of millions of noisy image–text pairs with contrastive alignment yields highly transferable representations that support strong zero-shot performance and robust cross-modal retrieval. These results underscore the importance of both data scale and contrastive objectives in shaping a well-structured joint embedding space. More recent LVLMs further extend this paradigm to broader multimodal inputs and increasingly complex downstream tasks, demonstrating that scaling model capacity, data set size, and modality diversity systematically enhances the quality of learned multimodal representations (Yuan et al., 2021; Alayrac et al., 2022; Chen et al., 2024a; Wang et al., 2023c). Together, attention-based cross-modal architectures and large-scale contrastive pretraining endow LMMs with superior abilities to capture inter-modal similarities, align semantic information, and generalize across heterogeneous data sources. These capabilities underscore the relevance of LMMs to industrial QC, where diverse data modalities must be interpreted in context to parameter designs, detect defects, trace root causes and generate actionable insights. By bridging the semantic gap across modalities, LMMs not only enhance decision accuracy but also lay the foundation for more autonomous, explainable and adaptive QC systems.

2.1.3 Few-shot generalization

Manufacturing quality data often exhibit severe class imbalance: defects and failures are rare, while normal operations are abundant. This creates a few-shot learning scenario where only a handful of examples (or even none) are available for certain fault modes (Liang et al., 2023). Classical few- and zero-shot techniques, such as metric-based meta-learning or prototype methods, can in principle address data scarcity, but they typically require carefully engineered architectures and task-specific training. Additionally, their transferability across heterogeneous products and modalities is limited. In contrast, LMs, including but not limited to LLMs, are pre-trained on web-scale multimodal data and learn rich and semantically structured representations that can be reused across tasks. This broad prior knowledge enables strong few-

and zero-shot performance via mechanisms such as feature reuse, semantic alignment and in-context learning, where a model conditions on a small number of examples or purely textual descriptions provided in the input rather than on task-specific fine-tuning (Brown et al., 2020; Radford et al., 2021, Wang et al., 2023b). For example, LVLM like CLIP can recognize novel defects from just a description and a few images, thanks to their extensive pre-training on diverse visual concepts (Radford et al., 2021). By leveraging knowledge from large scale pre-training, LMs use the input itself as a guide for adaptation. Unlike traditional models, which rely on large labeled data sets and fine-tuning, models like GPT-3 and CLIP can rapidly adapt to new tasks with just a few examples or prompts. Moreover, since LMs learn high-level features, they are less prone to overfitting small data sets and can generalize more effectively under data imbalance. By exploiting the inherent few- and zero-shot capabilities of LMs, AI-driven QC becomes viable even for rare defects, new product introductions and rapidly changing manufacturing environments, where relying on conventional few-shot techniques or collecting large task-specific data sets is uneconomically and operationally infeasible.

2.2 Task adaptability

QC in manufacturing is inherently multi-faceted. Practitioners must monitor processes in real-time, diagnose root causes and predict future quality outcomes. Each of these tasks entails distinct demands regarding data representation, modeling strategies and interpretability, which collectively pose significant challenges for traditional ML approaches when confronted with the inherent complexity and variability of industrial data. Instead of training separate models for each task, multi-task learning (MTL) provides a unified framework in which a single model learns multiple tasks simultaneously, enabling shared representations and cross-task knowledge transfer (Zhang and Yang, 2022; Liu et al., 2019; Caruana, 1997). LMs are inherently well-aligned with the principles of MTL due to several training and architectural properties: Their core strengths lie in general-purpose representation learning, enabled by pre-training on massive and diverse corpora that span diverse functions, domains and task types. This broad exposure equips LMs with highly transferable internal representations, reducing the need for task-specific modeling and manual feature engineering (Liu et al., 2020). Furthermore, Transformer-based LMs typically operate under a unified sequence-to-sequence or encoder-decoder architecture, allowing them to accommodate a wide variety of input-output formats with minimal structural modifications. This architectural uniformity supports flexible integration of tasks using either shared heads or prompt-based conditioning, making them well-suited for scalable multi-task adaptation (Chen et al., 2024b).

Recent advances across multiple domains have demonstrated the effectiveness of LMs as a foundation for MTL. A representative example in NLP is the text-to-text transfer transformer (T5) model (Raffel et al., 2020), which reformulates all tasks ranging from translation and summarization to question answering into a unified text-to-text paradigm. This task-agnostic formulation supports large-scale multi-task pretraining, allowing the model to learn shared semantic representations across tasks and achieve state-of-the-art results on numerous benchmarks. However, it also illustrates the importance of task compatibility, as performance may degrade when mixing highly dissimilar tasks without careful balancing strategies.

The success of unified LMs in NLP has inspired similar approaches in other modalities. In CV, the vision transformer (ViT) (Dosovitskiy et al., 2020) exemplifies this trend by treating images as sequences of patches, enabling the use of self-attention to learn general-purpose visual features. ViT has shown strong multi-task capability across classification, detection and segmentation tasks, reducing the reliance on task-specific architectures. Extending ViT, the swin-transformer (Liu et al., 2021) introduces a hierarchical structure with shifted window attention, improving the modeling of both local and global visual features. This allows for scalable MLT while maintaining high computational efficiency, making it suitable as a general-purpose backbone for dense vision tasks. In time series analysis, the UniTS framework (Gao et al., 2024) adapts the Transformer architecture to jointly model forecasting, classification, anomaly detection and imputation within a single network. By capturing universal representations of temporal dynamics, UniTS demonstrates the value of shared learning even in sequential domains with heterogeneous targets.

The versatility of LMs in supporting MTL extends to medical applications as well, where multiple related prediction tasks often coexist, for example, diagnosing several related conditions, or simultaneously predicting a disease and its expected progression. Kim et al. (2023) develop an MTL framework to predict multiple chronic diseases simultaneously, leveraging the interdependence of disease states to improve predictive accuracy. El-Sappagh et al. (2020) integrate multimodal clinical data to jointly solve classification and regression tasks related to Alzheimer's disease, while Yang et al. (2020) propose a multi-task model for lung cancer detection along with auxiliary respiratory disease diagnoses. To manage task imbalance and prevent overfitting, their design introduce periodic focusing and internal-transfer weighting strategies, techniques that improve robustness by modulating learning focus across tasks. In autonomous driving and robotics domain, MTL enables comprehensive perception and decision-making by combining various subtasks. The M3Net framework (Chen et al., 2025), for instance,

integrates object detection, semantic segmentation and 3D occupancy prediction into a single network through a query-token interaction mechanism. This unified approach not only achieves superior performance compared to separate task-specific models but also benefits from cross-task feature reinforcement where informative features for one task improve performance in others.

Pre-trained on massive and heterogeneous data sets, LMs have demonstrated strong generalization across domains when applied in MTL settings. Their inherent support for task integration, modular architecture and capacity to capture rich contextual representations make them well-suited as the core of unified and scalable QC systems. By facilitating shared understanding across tasks, such models not only enhance consistency and knowledge transfer but also substantially reduce deployment complexity.

2.3 Augmented intelligence

In modern QC systems, it is no longer sufficient to make accurate decisions, the ability to explain those decisions and interact naturally with human operators has become equally essential. In industrial anomaly detection, a recent survey emphasizes that a key objective of LMs is to generate interpretable detection results that describe anomalies in terms of color, shape and category (Yang et al., 2025), bridging the gap between automated decisions and human understanding. Traditional models often fall short in providing interpretability and seamless human-machine communication. LLMs such as ChatGPT (Ouyang et al., 2022) and DeepSeek (Bi et al., 2024) offer a compelling alternative with the remarkable capabilities in comprehending (Peng et al., 2024), extracting (Xu et al., 2024) and reasoning (Wu et al., 2024). Leveraging their advanced NLP capabilities, LLMs can generate context-aware explanations and engage in dynamic and interactive dialogs, bridging the gap between automated decision-making and human understanding.

The generative capabilities of LLMs have demonstrated remarkable effectiveness across a wide range of application domains. One of the most notable areas of success is software development, where LLMs have shown strong potential in automating and enhancing programming-related tasks. Models like OpenAI's Codex (Chen et al., 2021), trained on extensive corpora of source code, can translate natural language descriptions into executable code across multiple programming languages with high syntactic and semantic accuracy. Beyond code synthesis, LLMs have also proven valuable in tasks such as debugging, code explanation and exploration of alternative implementations (Sun et al., 2022). These capabilities underscore LLMs' strengths in logical reasoning, error identification and iterative refinement, skills that are not only vital in software engineering, but also directly transferable to the diagnosis and resolution of quality issues in industrial systems.

In the medical and healthcare domain, LLMs have made remarkable strides, tackling tasks that require expert-level knowledge and complex reasoning. A prominent example is the medicine-pathways language model (Med-PaLM) (Singhal et al., 2023), the first AI system to surpass the passing threshold on US medical licensing examination style questions. This milestone illustrates the capacity of LLMs to encode extensive domain-specific knowledge and to reason effectively over complex medical content. Building upon this foundation, Med-PaLM 2 (Singhal et al., 2025) introduces further innovations by addressing the challenges of long-form medical question answering and integration into real-world clinical workflows. This is accomplished through a combination of enhanced base model capabilities, targeted fine-tuning on medical data and advanced techniques such as ensemble refinement (Madaan et al., 2023) and retrieval-augmented reasoning (RAG) (Lewis et al., 2020). These strategies significantly improve the model's factual grounding and reasoning accuracy. In human evaluations, physicians prefer the responses generated by Med-PaLM 2 over those of generalist doctors on most clinical questions and rate its advice as comparably safe. These achievements highlight LLMs' ability to apply expert knowledge and reason with justification using natural language.

Similarly, in enterprise knowledge management, LLMs are transforming how organizations retrieve and utilize vast amounts of unstructured data with unprecedented efficiency. Through natural language understanding and generation, they facilitate retrieval and synthesis of information from documents, manuals and conversational logs. Jiang et al. (2024) propose an LLM-based retrieval-generation framework tailored for enterprise knowledge bases named as EKGR, addressing challenges such as limited annotations and data privacy. It combines instruction-tuned question generation, a relevance-aware teacher-student retriever training strategy and chain-of-thought fine-tuning for answer generation, achieving strong performance on real-world data sets with minimal supervision. Indeed, there is growing recognition that LLMs may serve as a new backbone for knowledge management in organizations (O'Leary, 2023; Lang et al., 2024). These successes in medical and enterprise contexts show that LLMs can act as knowledge integrators and conversational agents, making them highly relevant to QC scenarios, where engineers often need to diagnose complex production issues, consult large volumes of historical failure reports and interpret best-practice guidelines, all tasks that benefit from expert-level reasoning and natural language interaction.

The cross-domain success of LLMs underscores their potential as general-purpose cognitive engines for industrial quality applications. By expressing insights in natural language, LLMs turn raw predictions into actionable, human-aligned explanations. Beyond interpretability,

LLMs enhance interaction by linking data-driven analysis with human communication. Their ability to understand and summarize unstructured sources enables the extraction and reuse of valuable domain knowledge. In smart manufacturing, engineers can query LLMs in natural language (e.g., "Why has the defect rate increased this week"?) and receive reasoned and evidence-based answers grounded in real-time data and historical trends. These capabilities position LLMs not as passive reporting tools, but as interactive and adaptive systems that integrate human expertise with machine reasoning. As their capabilities advance and industrial adoption accelerates, LLMs are poised to become essential to intelligent QC.

3 Enabling large models for quality control

Building on the foundational capabilities of LMs, recent research has begun to explore their concrete applications in industrial QC, evolving from conceptual exploration to practical deployment. This section reviews recent advances that leverage LMs to address core challenges in QC, including multimodal perception for integrating heterogeneous data, multitask reasoning for unified process understanding, few-shot and zero-shot learning to mitigate data scarcity and generative AI to enable interactive and explainable quality management. These developments collectively demonstrate the expanding role of LMs in advancing intelligent and robust QC systems.

3.1 Multimodal perception

Multimodal perception plays a critical role in industrial quality inspection, where integrating diverse data sources, such as visual images, acoustic signals, vibration data and thermal measurements, enable more accurate and robust defect detection. An overview of commonly used sensing modalities, along with their feature representations, is provided in Table 1. By leveraging complementary information from different modalities, LMs can capture more comprehensive process signatures that single-modality approaches often overlook. A range of multimodal fusion approaches developed to facilitate this integration is summarized in Table 2.

3.1.1 Attention-based multimodal fusion

Although attention mechanisms are not exclusive to LMs, their flexibility and strong capacity for relational modeling have made attention-based fusion a predominant architectural choice in contemporary LMMs. This trend has also carried over to industrial QC applications, where such models benefit from Transformer-based LMs' ability to capture long-range dependencies in sequential data given that most industrial quality signals are inherently time-series in nature. At the same time, Attention enables dynamic modeling of both intra- and inter-modal dependencies, effectively capturing temporal patterns in sensor sequences and fine-grained cross-modal correlations among heterogeneous data sources (Zhao et al., 2024). Specifically, self-attention layers allow each modality to model its internal temporal or spatial dependencies, while cross-attention facilitates interaction across modalities, enabling the model to align sensor trajectories with corresponding video frames or highlight visual regions associated with anomalous process behaviors. This fine-grained and context-adaptive interaction is particularly advantageous in noisy industrial environments because attention itself inherently suppresses irrelevant features and amplifies modality-specific cues that are most informative for

Table 1 Overview of sensing, odalities, data types and feature representations in industrial quality inspection

Modality type	Data type	Feature representation
Process	PLC logs, actuator positions, process variables, structured signals	Time-series embeddings, attention-based features
Image	RGB, high-resolution images	CNN, ViT embeddings
Acoustic	Microphones, ultrasonic sensors	Spectrograms, time-domain features
Vibration	Accelerometers, vibration sensors	FFT spectra, envelope signals
Thermal	Infrared images, thermal sensors	Thermal distribution maps
Point cloud	LiDAR, structured light	Voxelized grids

Table 2 Multimodal fusion techniques for quality control

Fusion method	Core mechanism	Learning paradigm
Attention-based fusion	Models intra-/inter-modal dependencies using self- and cross-attention mechanisms	Supervised (Costanzino et al., 2024; Guo et al., 2022; Li et al., 2024c; Nagrani et al., 2021; Wu et al., 2025b; Jiao et al., 2024; Zhao et al., 2024)
Domain knowledge-guided alignment	Injects expert rules, physics priors, fault ontologies and standards into LMMs to guide multimodal alignment.	LoRA fine-tuning + Domain prompts (Bianchini et al., 2025)
Contrastive fusion	CLIP-style shared embedding space using contrastive objectives.	Unsupervised (McKinney et al.,2025; Wang et al., 2025)

the task (Guo et al., 2022; Jiao et al., 2024;Nagrani et al., 2021). Compared to traditional multimodal fusion techniques, attention-based fusion provides deep and token-level interaction between the features of each modality (Baltrušaitis et al., 2019). This enables the model to learn fine-grained dependencies across different inputs, making it especially effective when modalities provide complementary information that needs to be fused dynamically.

However, applying attention-based fusion model to industrial QC still requires tailored model design to address the specific data types and scenarios encountered in this domain. For instance, Li et al. (2024c) propose an industrial fault diagnosis system that incorporates an attention-guided fusion module to jointly analyze video streams and time-series process data. This enables precise diagnosis even under highly complex production scenarios with dynamic operational conditions. Wu et al. (2025b) further extend this paradigm by introducing a cross-modal Transformer for anomaly detection in magnesium smelting, utilizing self-attention to capture intra-modal features and bidirectional cross-attention to model inter-modal correlations between video data and process variables. Costanzino et al. (2024) introduced a cross-modal attention framework inspired by transformers to align and fuse features from 3D point clouds and 2D images. By dynamically capturing correlations across spatial and structural domains, their methods excel in detecting subtle anomalies that are often invisible in purely visual inspection. Beyond detection and localization tasks, Li et al. (2024e) also explore attention-based fusion in the context of quality prediction. Their work introduces a multimodal quality prediction algorithm that integrates time-series sensor data, image-based inspection data, and process parameters using an anomalous energy tracking attention mechanism. By dynamically modeling both temporal dependencies and cross-modal interactions, their approach significantly improves prediction accuracy and robustness under complex and noisy manufacturing conditions.

3.1.2 Domain knowledge-guided multimodal alignment

While LMMs excel at learning generalized representations across diverse modalities, their direct application to specialized industrial tasks can remain limited due to the lack of embedded expert knowledge. Bridging this gap is essential to unlock the full potential of LMMs in industrial settings. In this context, domain knowledge-guided multimodal alignment refers to approaches that inject structured domain priors, such as physics-based rules, fault-mode ontologies, or expert-defined labeling heuristics, into the alignment process so that representations across modalities become consistent with the way human experts conceptualize diagnostic relationships. Such guidance may be incorporated at multiple levels, including rule-based construction of multimodal training pairs, knowledge-conditioned contrastive learning objectives, or architectural modules that explicitly encode relations defined in domain knowledge graphs or rule bases. Recent surveys on multimodal alignment and fusion emphasize that these knowledge-aware mechanisms are critical for achieving robust performance in domains where raw data alone are insufficient to capture all relevant semantic structure (Li and Tang, 2024).

A representative example of domain knowledge-guided alignment for industrial QC is VSLLaVA, a large multimodal foundation-model pipeline tailored for vibration signal analysis (Li et al., 2024f). Rather than relying solely on raw signal–text co-occurrence, VSLLaVA begins by using an expert rule-assisted signal generator to synthesize a large Signal–Question–Answer (SQA) corpus, in which each triplet links a vibration waveform to a diagnostic query and an expert-consistent explanation. The generator encodes core vibration-analysis concepts, including amplitude and frequency modulation patterns, harmonic structures, and characteristic bearing fault signatures, ensuring that the constructed SQA pairs faithfully reflect underlying physical mechanisms and established diagnostic rules. These triplets are then used to conduct instruction tuning via low-rank adaptation (LoRA) on a CLIP-based vision encoder and an LLM, thereby injecting domain-specific notions of normal behavior, incipient faults, and severe degradation into the shared embedding space. Finally, VSLLaVA introduces a knowledge-aware evaluation loop in which an LLM, guided explicitly by expert rules, evaluates the correctness and relevance of generated outputs for tasks such as parameter identification, waveform interpretation, and fault diagnosis. Complementary to this, Bianchini et al. (2025) propose a multimodal RAG framework for quality anomaly detection in industrial electronics manufacturing. Their approach leverages domain knowledge in the form of engineering manuals and inspection standards to align structured measurements and image-based detection outputs with textual diagnostic rules. By integrating document-grounded retrieval with LLM-based interpretation, their system demonstrates how symbolic domain expertise can guide the alignment of multimodal inputs in real-world QC pipelines.

3.1.3 Multimodal input-driven unsupervised learning

The inherent ability of LMMs to align heterogeneous modalities into a shared semantic space significantly alleviates the challenges of data scarcity, thereby enabling robust unsupervised learning in QC tasks. Recent studies have demonstrated that, even without labeled defect data, LMM-based multimodal learning frameworks can learn modality-specific encoders whose embeddings are aligned through contrastive or reconstruction-based

objectives defined purely over synchronized inputs. Once trained, the resulting joint embedding space can support zero- or few-shot tasks.

A representative example is the work by McKinney et al. (2025), which proposes an unsupervised multimodal fusion framework inspired by CLIP for industrial process monitoring. Their model constructs modality-specific encoders for the distinct sensors and employs contrastive learning to align representations across modalities without any labeled defect data. By projecting these modalities into a common embedding space, the model captures cross-modal patterns that are highly indicative of quality anomalies, facilitating defect detection in a fully label-free manner. Extending multimodal unsupervised learning to 3D quality inspection, recent advances have adapted CLIP-like vision–language alignment to point cloud data. Zuo et al. (2024) propose CLIP3D-AD, which projects each 3D object into a set of multi-view rendered images and fuses 2D features through adapter-based fine-tuning and coarse-to-fine decoding. This architecture effectively bridges the modality gap between 2D visual representations and 3D geometric structures, enabling the model to perform few-shot and zero-shot anomaly detection without memory banks or extensive labels, and significantly outperforming state-of-the-art industrial 3D inspection methods. Building on the same motivation, Zhu et al. (2023) introduce PointCLIP V2, a hybrid multimodal 3D anomaly detection framework that unifies CLIP and GPT-3 as a single 3D open-world learner by prompting CLIP's visual encoder with a realistic shape projection module that produces high-quality depth maps from point clouds and prompting GPT-3 to generate 3D-specific text as the input to CLIP's textual encoder, thereby narrowing the domain gap between 3D data and pre-trained vision–language knowledge and enabling zero-shot/few-shot 3D understanding. Wang et al. (2025) propose M3DM-NR, an unsupervised hybrid multimodal framework using pre-trained CLIP and Point-BIND models. By aligning RGB and point cloud modalities through contrastive learning and applying multi-scale aggregation techniques, M3DM-NR improves anomaly detection accuracy, making it highly effective in noisy real-world industrial environments.

Taken together, these three research directions reflect complementary explorations that demonstrate the value of LMMs for QC. Although they approach the problem from different angles, they collectively show how multimodal information can be effectively integrated, structured, or learned without extensive annotation. Overall, these developments suggest that LMMs are evolving into unified perceptual systems capable of providing scalable, accurate, and generalizable quality assessments through a comprehensive and context-aware understanding of complex manufacturing processes.

3.2 Multi-task reasoning

In recent QC systems, Transformer- and attention-based architectures often serve as LM-style backbones and provide strong shared representation learning. Compared with general-purpose MTL settings, MTL in industrial QC is substantially more complex. Tasks such as quality prediction, process monitoring, fault diagnosis and health prognostic are tightly linked through temporal dependencies and cross-stage causal relationships inherent to manufacturing processes. Industrial systems also place higher demands on interpretability to support decision-making and root-cause analysis. As summarized in Table 3, these characteristics require models that can learn rich and unified representations while also using specialized mechanisms to regulate interactions among tasks in a stable and transparent manner. Such mechanisms include loss weighting, gradient balancing, multi-stage modeling and mixture-of-experts (MoE) structures, which help control how tasks influence each other during training and inference.

3.2.1 Shared representation learning

One of the fundamental questions in MTL is how to share representations. In the context of LMs, shared representation learning typically involves building a common feature extractor which feeds into task-specific output heads. This structure allows the model to jointly leverage information across tasks. Han et al. (2025) formulate a CNN–Transformer hybrid model with a shared feature extractor and task-specific branches to simultaneously diagnose multiple bearing fault modes. This Transformer-enhanced architecture effectively captures both local signal patterns and long-range dependencies, yielding

Table 3 Key challenges and strategies in multi-task learning for industrial quality control

Challenge	Cause	Strategy	Method
Data imbalance	Few labels for some tasks	Attention, low-shot sharing	Xie et al. (2022); Han et al., (2025)
Shared representation	Task-general vs. task-specific	Hybrid encoders, split decoders	Cao et al. (2024); Han et al. (2025);Liu et al. (2023b);Wang et al. (2024b);
Multi-stage modeling	Cross-stage variable shifts	Latent state, stage regularization	DMMTL (Yan et al., 2021),
Task interference	Cross-task conflicts	Gradient balance, MoE, loss tuning	AWAMTL (Wu et al., 2025a), BMoE (Huang et al., 2023)
Task specialization	Mixed correlation across tasks	MoE routing, gated experts	AMSF-AMoE (Zhou and Wang, 2024), BMoE (Huang et al., 2023)

robust fault recognition even under heavy noise and data imbalance. Similarly, Cao et al. (2024) address MTL by proposing an enhanced hybrid LSTM–Transformer model that jointly learns shared representations and generates task-specific predictions in real time for multiple related targets. By integrating LSTM-based sequence modeling with self-attention, the framework supports continuous adaptation through online learning and enables efficient deployment via knowledge distillation while maintaining strong multi-task predictive performance. Xie et al. (2022) further introduce an attention-guided multi-task network that simultaneously identifies fault types and severity levels, rather than one at a time, enabling the model to share global features and improve fault diagnosis under limited data conditions. Wang et al. (2024b) propose a target-related Transformer model that learns a shared representation space guided by specific quality objectives, allowing effective modeling of multiple quality indicators within a single framework. To enhance model interpretability, Liu et al. (2023b) propose DMRI-Former built on a shared encoder–decoder Transformer structure, and they further design a data-mode-related interpretable self-attention that first clusters samples into data modes and then performs homomode attention with optional cross-mode attention, so that the learned attention weights can be visualized as heatmaps for key-sample analysis in multimode industrial process prediction. Liu et al. (2024) extend the shared representation paradigm to time series with Multirate-Former, a hierarchical architecture that models multistep quality prediction across processes with heterogeneous sampling rates, leveraging a common representation across temporal scales.

3.2.2 Dynamic task balancing

Building on the deep exploration of shared representations, an equally important direction in LM-based industrial MTL is how to regulate and balance the learning dynamics among multiple tasks, especially when tasks have heterogeneous scales or difficulty. In industrial QC, where tasks can range from predicting continuous metrics (tool wear, remaining useful lifetime, quality scores) to classifying discrete states (healthy/faulty, defect types), such adaptive balancing methods are invaluable. Yan et al. (2021) propose a deep multistage MTL framework tailored for quality prediction in complex multistage manufacturing systems, where numerous correlated quality indices are measured across sequential stages. The framework dynamically balances tasks by learning latent state representations and introduces stage-wise regularization, specifically group lasso penalties and robust Huber loss, to adaptively suppress uninformative outputs and select relevant inputs. This design enables end-to-end joint modeling of all quality variables while effectively mitigating task interference and enhancing interpretability through structured sparsity and gradient-based diagnostics. Wu et al. (2025a) propose an attention-based weight adaptive multi-task learning framework to address slab head shape defects in the hot rough rolling process, a critical quality issue characterized by nonlinear and multivariate deformation. Task weights are adaptively updated at each training iteration based on gradient disparities and task-specific convergence rates, ensuring balanced optimization despite differing learning dynamics. This gradient-aware adjustment mechanism enhances robustness and accuracy by mitigating task interference throughout the training process. Moreover, MoE-based architectures (Jacobs et al., 1991) have overcome negative transfer in diverse multi-task QC by dynamically routing inputs to specialized experts, balancing shared learning with task-specific adaptability across varying manufacturing conditions. For instance, Zhou and Wang (2024) develop an adaptive multi-scale feature fusion and adaptive MoE multi-task model for concurrent health status assessment and remaining useful lifetime prediction. This design effectively captures inter-task dependencies while addressing heterogeneity, significantly enhancing model generalization on complex data sets like engine degradation and tool wear. Additionally, Huang et al. (2023) present a balanced MoE to estimate multiple quality variables, which consists of a multi-gate MoE module to portray task relationships and a task gradient balancing module to balance the gradients among tasks dynamically. Both of them cooperate to mitigate the negative transfer problem.

In summary, LM-style Transformer backbones play a central role in industrial multi-task QC by providing unified, expressive and context-aware representations that capture cross-task dependencies, temporal dynamics and multimodal process variations. Building on this representational foundation, MTL algorithms further tailor the learning process to the specific demands of complex manufacturing environments. The progress observed across recent studies reflects the synergy between these two components: strong LM-driven shared representation learning and targeted MTL-driven task interaction modeling. Their integration offers a promising direction for future research, enabling multi-task QC systems that are not only more accurate and robust but also more aligned with real-world industrial requirements.

3.3 Few/zero-shot reasoning

Supervised QC methods heavily depend on large amounts of labeled defect data, which is costly and impractical given the rarity and diversity of real-world defects (Zajec et al., 2024). This limitation underscores the importance of few- and zero-shot learning, enabling models to generalize to unseen categories with minimal or no annotations. Characterized by strong cross-domain generalization, multimodal alignment and prompt-based adaptability, LMs offer a scalable solution, facilitating label-efficient and flexible inspection. A categorization of such

LM-powered few- and zero- shot strategies is provided in Tables 4 and 5.

3.3.1 Few-shot learning

LMs exhibit remarkable few-shot capabilities in QC tasks, primarily due to their large-scale pretraining on diverse visual, textual, or sensor data, which enables the learning of highly transferable and generalized representations. For instance, Megahed et al. (2025) demonstrated that CLIP's vision–language pretraining equips it with robust visual features that transfer effectively to surface defect classification and 3D-print anomaly detection with as few as 10–100 labeled samples per class. To further bridge this gap, researchers have infused domain knowledge via prompting. Jin et al. (2024a) propose the SEM-CLIP, which guides CLIP's attention to semiconductor wafer defect regions using expert text prompts. By customizing CLIP's vision–language alignment, their method accurately classifies and segments wafer defects with minimal labeled examples. Beyond vision, Eldele et al. (2025) introduce UniFault, a transformer-based foundation model pretrained on over 6.9 million industrial sensor records, achieving state-of-the-art fault detection in predictive maintenance with minimal samples from new machines, showcasing the cross-domain generalization power of foundation models. Similarly, LLMs leverage their in-context learning to solve manufacturing few-shot problems cast into language-like formats. Fang et al. (2024) successfully reformulate 3D-print defect detection as a text-prompted task, enabling a GPT-style model to achieve 96.2% accuracy with only a few image–caption pairs and expert explanations, without any retraining. Complementing these approaches, Hu et al. (2024) employ generative foundation models to address data scarcity by proposing AnomalyDiffusion, which synthesizes diverse anomaly–mask pairs from limited examples, enhancing the performance of downstream defect detection models. These works collectively demonstrate that the few-shot success of LMs in industrial quality tasks stems from their broadly learned semantic priors and strong capacity for rapid task adaptation via prompting or in-context learning, even under significant domain shifts.

3.3.2 Zero-shot learning

Recent advances demonstrate that LMs possess remarkable zero-shot capabilities. However, as Zhong et al. (2021) show in the NLP domain, zero-shot performance is determined not merely by model scale but also critically depends on how models are trained and adapted, including the choice of meta-tuning objectives and prompt formulations. This implies that directly reusing the foundation model or applying naïve fine-tuning is rarely sufficient for industrial zero-shot QC, where defects are subtle, data are scarce, and task specifications are highly structured. Instead, effective zero-shot QC relies on task-aligned prompt engineering, tailored input representations, and lightweight adaptation mechanisms.

Within industrial QC, one prominent line of work evaluates and adapts the segment anything model (SAM) as a universal segmentation backbone. Introduced by Kirillov et al. (2023), SAM is a promptable segmentation foundation model trained on over one billion masks and exhibits strong zero-shot transfer on natural-image benchmarks. Motivated by this potential, Song et al. (2024) systemati-

Table 4 Categorized strategies for few-shot learning in industrial quality control

Method category	Adaptation strategy	Domain knowledge	Modality	Strength	Limitation
Pretrained vision-language	Prompt-based inference	Minimal	Vision + Text	General, scalable, data-efficient	Lacks domain specificity (Megahed et al., 2025)
Prompt-guided fine-tuning	Prompt or attention tuning	Expert prompts	Vision + Text	Better localization, data-efficient	Prompt quality sensitive (Jin et al., 2024a)
Sensor foundation models	Few-shot tuning	Embedded in training	Time-series, vibration, multi-sensor	Cross-device transfer	Needs large pretraining data (Eldele et al., 2025)
Language reformulation	Text-format inference	Caption prompts	Vision → Text	High generalization, no retraining	Needs precise phrasing (Fang et al., 2024)
Generative augmentation	Synthetic data generation	No explicit knowledge	Vision	Data boost, helps detection	Extra cost, variable quality (Hu et al., 2024)

Table 5 Categorized Strategies for zero-shot learning in industrial quality control

Methodcategory	Mechanism	Modality	Strength	Limitation
Template-based language prompting	General prompts (e.g., CLIP)	Vision	Simple, generalizable	Weak in noisy/fine tasks (Zhong et al., 2021)
Universal segmentation models	Pretrained models (e.g., SAM)	Vision	Label-free, broad use	Low industrial specificity (Kirillov et al., 2023; Song et al., 2024)
Unsupervised prompt composition	Auto-generated prompts	Vision + Rule templates	Fully unsupervised	Unstable prompt quality (Era et al., 2024; Huang et al., 2025)
Hybrid domain- prompting	Text + symbolic input	Vision + Text + Symbols	Precise, interpretable	Complex setup (Cao et al., 2025)

cally evaluate SAM on industrial defect data sets, such as strip-steel, tile, and rail surfaces, under various prompting modes and show that, despite its impressive generalization on natural images, out-of-the-box SAM consistently underperforms well-tuned saliency-based and defect-specific segmentation methods on low-contrast, fine-grained industrial surfaces. This result clarifies that SAM's generic priors are not directly sufficient for reliable QC: the strong pre-trained segmentation ability must be coupled with task-aware prompting, domain adaptation or both. To mitigate this gap, subsequent studies explore automated or scene-level prompting strategies. Era et al. (2024) generate contextual point prompts through unsupervised feature clustering and use them to guide SAM for porosity segmentation in additive manufacturing without manual labels. Huang et al. (2025) further propose SAID, which builds a global scene embedding from a few annotated examples and uses it as a high-level prompt to enable one-shot segmentation in new scenes. These efforts trace a clear trajectory: from direct SAM reuse to structured and scene-aware prompting tailored to industrial environments.

Beyond segmentation models, vision–language approaches extend zero-shot QC to both detection and understanding. Jeong et al. (2023) propose WinCLIP, which mitigates CLIP's weak localization of small defects by extracting overlapping window features and aggregating them at patch and image levels. WinCLIP also employs a compositional prompt ensemble combining multiple state descriptors (e.g., normal, anomalous) and template variations, allowing it to better capture the linguistic diversity of defect descriptions. By comparing each patch's visual embedding with corresponding text embeddings, WinCLIP identifies anomalous regions without any fine-tuning and achieves superior zero-/one-shot performance on industrial anomaly benchmarks. Cao et al. (2025) propose AnomalyVLM, a framework that moves beyond reference-based anomaly detection by leveraging product standards and engineering specifications as a structured source of prior knowledge. Instead of relying on clean exemplars, AnomalyVLM parses these documents to construct hybrid prompts that jointly encode: (i) textual descriptions of nominal appearance and typical defect patterns, (ii) symbolic rules and region-level constraints specified in the standards, and (iii) region indices that anchor the prompts to specific parts of the product. These hybrid prompts are then used to personalize a pretrained vision–language model, enabling it to reason about what constitutes normal versus anomalous conditions under the guidance of domain knowledge. Evaluations on multiple data sets show that AnomalyVLM achieves state-of-the-art zero-shot performance, demonstrating the effectiveness of coupling foundation models with explicit, structured prior knowledge for industrial QC.

Across all these examples, the common thread is that large-scale pretraining provides a powerful prior: whether it is vision–language alignment, universal segmentation, temporal dynamics, or linguistic reasoning. While their unsupervised learning and generative abilities can effectively reduce reliance on labeled data, such as by synthesizing high-quality anomalies for data augmentation, the complexity of real-world industrial tasks still demands the integration of domain expertise through prompt engineering, fine-tuning, or hybrid frameworks to ensure alignment with specific operational requirements.

3.4 Generative AI Interaction

Recent advances have demonstrated the potential of LLMs to serve as interpretable and intelligent interfaces in industrial environments. Their ability to process unstructured data makes them effective in extracting insights from quality-related information (Chkirbene et al., 2024), while also serving as a bridge between human expertise and automated systems. The core functions of LLMs in this context are summarized in Table 6.

3.4.1 LLMs for explainable human–machine interaction

As manufacturing systems become increasingly complex, the demand for intuitive and explainable interfaces between humans and machines is more critical (Moosavi et al., 2024). Recent surveys on smart and human-centric manufacturing argue that future QC solutions must not only detect defects but also communicate their decisions in a way that aligns with operators' mental models and process knowledge (Abhilash et al., 2024). LLMs are well suited to this role: through instruction tuning and large-scale pre-training on code, technical text and natural language, they can translate low-level machine instructions, sensor logs, or images into coherent explanations

Table 6 Core roles of LLMs in industrial quality control

Application area	Main function	Knowledge requirement	Industrial benefit	Typical scenario
Explainable interaction	Translate machine code; Data to dialog	Low	Enhance interpretability, usability, and operator trust	Code explanation (Jignasu et al., 2023); quality guidance (Badini et al., 2023)
Domain-aware reasoning	Causal diagnosis, retrieve over expert knowledge bases	High	Enable automated fault analysis	CausalKGPT (Zhou et al., 2024); SOP query (Kernan Freire et al., 2024)
Multi-agent collaboration	Coordinate QC tasks via dialogs	Medium	Support autonomous, real-time QC	Agent-driven process control (Lim et al., 2024); Prognostics co-pilot agents for maintenance (Lukens et al., 2024)

and interactive dialogs. This capability enables operators to better understand, monitor and control automated processes, fostering more transparent and effective human-machine collaboration.

One stream of work focuses on code-centric interaction in additive manufacturing, where QC often hinges on whether process parameters and G-code are appropriate. Jignasu et al. (2023) evaluate several LLMs for G-code debugging, manipulation and comprehension, designing prompts that ask the model to detect syntax or logic errors, propose parameter changes, and explain the impact on part geometry and print quality. Badini et al. (2023) go further by using ChatGPT not only to diagnose typical printing issues, such as warping, stringing or bed detachment, but also to recommend specific adjustments to temperature, speed and flow settings, showing that non-expert users can obtain expert-level troubleshooting guidance through dialog. A second stream leverages multimodal to make visual QC more transparent. In industrial anomaly detection, AnomalyGPT (Gu et al., 2024) combines an LVLM with simulated anomaly: description pairs and a learned prompt module so that the model can both localize defects in images and describe them in natural language. At inference time, inspectors can interact with the system via multi-turn dialog, (for example, asking what type of defect is present, how severe it is, or which region requires rework,) while the model updates its responses based on both the image and the conversation context. Compared with code-centric approaches, which operate on symbolic process representations, AnomalyGPT directly interfaces with visual evidence from QC and emphasizes interactive explanation over scalar anomaly scores, eliminating manual threshold tuning and making defect decisions easier to scrutinize and trust.

3.4.2 LLMs powered by domain expertise

While general-purpose LLMs offer impressive linguistic capabilities, their value for industrial QC becomes most evident when they are explicitly enriched with domain-specific knowledge. Zhou et al. (2024) develop CausalKGPT, which integrates a causal quality knowledge graph built from defect surveys, inspection reports, and maintenance records. The model is fine-tuned with prompts grounded in this graph, enabling it to interpret defect descriptions and generate root-cause explanations aligned with real manufacturing causal structures. Evaluations in aerospace machining show that CausalKGPT provides more accurate and engineering-relevant diagnoses than generic LLMs, illustrating how causal, graph-structured knowledge strengthens QC-related reasoning. A complementary line of work focuses on operational knowledge access. Kernan Freire et al. (2024) implement a retrieval-augmented LLM assistant trained on factory manuals, procedures, and diagnostic records. When operators pose natural-language questions, the system retrieves relevant documents and synthesizes grounded responses. Field studies indicate that it accelerates troubleshooting while allowing workers to contribute new insights that incrementally update the knowledge base. This human-in-the-loop RAG approach showcases how LLMs can democratize expert knowledge across production teams. Similarly, Wang et al. (2024a) propose a framework for empowering ChatGPT-like LLMs with localized knowledge bases tailored for prognostics and health management. By coupling general-purpose LLMs with structured plant-specific knowledge, including degradation modes, equipment hierarchies, and domain-specific diagnostic flows, the system improves interpretability and domain alignment for tasks such as fault analysis, inspection guidance, and maintenance decision support. Their approach reinforces the value of combining RAG-based knowledge injection and natural-language reasoning to make LLMs more context-aware, trustworthy, and usable in real-world industrial environments.

3.4.3 Multi-agent LLM for collaborative QC

Beyond single-agent systems, researchers have begun designing multi-agent frameworks that integrate multiple generative models for collaborative QC. Lim et al. (2024) demonstrate a manufacturing multi-agent system in which each agent implemented using an LLM such as GPT-4 performs a specialized role (e.g., planning, inspection, troubleshooting) and communicates through natural-language messages. Within this architecture, agents articulate their reasoning, negotiate task assignments, and coordinate corrective actions, enabling collective decision-making that resembles a virtual team of quality engineers. Similarly, Lukens et al. (2024) propose a multi-agent LLM framework for prognostics and health management copilots. Their workflow decomposes PHM support into specialized agents for observation extraction, failure-mode extraction, recommendation generation, and recommendation evaluation, with structured function-calling outputs. Using a case-study data set of industrial PHM narratives, they show that RAG can improve the coverage and earlier identification of failure modes in recommended troubleshooting plans. The study also highlights limitations of LLM-based evaluation, indicating the need for more reliable evaluation protocols and domain grounding. Although these architectures are still in their early stages and often validated in simulated settings, they outline a compelling direction: networks of generative agents that can jointly monitor, diagnose, and support timely interventions for quality issues. Compared with single-agent systems, multi-agent designs naturally promote specialization, distributed reasoning, and coordinated interaction, making it plausible that interconnected agents could deliver continuous and adaptive QC assistance across the production workflow.

Together, these developments illustrate the expanding

role of generative AI in industrial QC. Whether through code interpretation, root-cause analysis, conversational support, or agent collaboration, LLMs are enabling more transparent, adaptive and accessible QC workflows across the production lifecycle.

4 Challenges and perspectives

In the previous section, we have discussed why and how the LMs can empower intelligent QC. This section attempts to discuss the challenges and prospects of these models from a broader and more global perspective.

4.1 Challenges

Despite the promising potential of LMs in QC, several critical challenges remain across multiple dimensions (see Fig. 3).

4.1.1 Data-level

For training and fine-tuning LMs, access to abundant high-quality data are a fundamental prerequisite. Beyond general challenges such as missing data and high noise, the data landscape in industrial QC has several distinctive characteristics that make applying LMs fundamentally different from other domains. First, modern manufacturing systems typically operate at very high yield, and thus most collected samples are defect-free, while real defective parts are scarce and often highly heterogeneous. Industrial anomaly detection studies consistently emphasize that benchmark data sets are built under a "many normal, few defective" regime, with unsupervised or one-class settings precisely because labeled anomaly samples are too rare and diverse for conventional supervised training. This lack of diverse training examples hinders model generalization, increasing the risk of overfitting and reducing effectiveness on new products or conditions. The scarcity of large-scale and high-quality data has become a recognized bottleneck, prompting efforts to collect compliant and diverse data sets.

Second, QC data are tightly constrained by data security and intellectual property considerations. Industrial data typically involve proprietary manufacturing processes and sensitive customer information, making companies reluctant to share raw data due to intellectual property and regulatory concerns. Review of federated learning in smart manufacturing and product lifecycle management highlight that operational data are distributed across different plants and business units, and that privacy and confidentiality concerns make centralized data pooling infeasible (Leng et al., 2025). For LMs deployed in QC, this implies that training and inference pipelines must account for both industrial data governance and LLM-specific leakage channels, which is much more restrictive than typical public-data settings.

Third, industrial QC is intrinsically multimodal, and the alignment of modalities is more complex than in standard image–text tasks. Visual inspection images, 3D scans, vibration and acoustic sensor signals, temperature and pressure traces, and structured quality reports all monitor different objects or subsystems in the production process. Multimodal benchmark further demonstrate that defects may only be visible in geometry or only in texture, requiring models to reason about physical structure rather than just semantic similarity (Bergmann et al., 2021). For LM-based QC assistants, this means that multimodal alignments must respect process physics and temporal causality, e.g., how a small change in a process variable manifests as a subtle surface defect, rather than the looser alignments typical of regular image–caption pairs.

Data-level challenges fundamentally limit the training and generalizability of LMs in industrial QC applications. To overcome them, high-quality, standardized data sets must be prioritized. This calls for closer collaboration between industry and academia to co-develop benchmark data sets that reflect real-world variability and task-specific requirements.

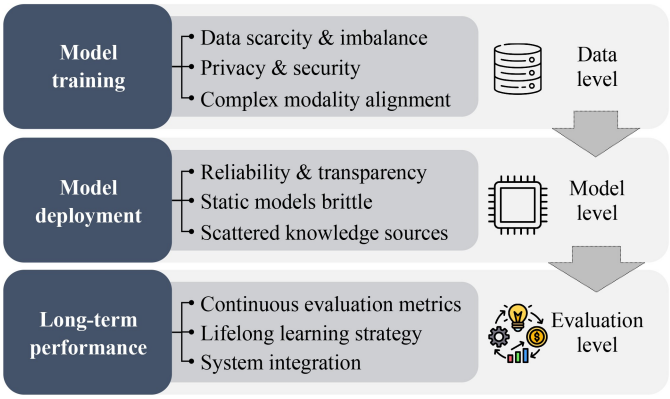

Fig. 3 Key challenges in leveraging large models for quality control.

4.1.2 Model-level

On the model side, the probabilistic nature of LMs conflicts with the strict reliability requirements of industrial QC. Even the most advanced LLMs are known to generate fluent but incorrect content, and recent surveys identify hallucination as a key barrier to deploying foundation models in safety-critical applications (Sahoo et al., 2024). For the reason that QC decisions directly affect product integrity and operational safety, LMs in QC must satisfy much tighter bounds on acceptable hallucination rates and must be coupled with strong grounding, verification, or retrieval-augmented mechanisms than in generic conversational applications. At the same time, interpretability is essential for industrial acceptance: engineers and supervisors must understand why a model flags defects or suggests parameter changes. As Li et al. (2024d) note, the integration of AI in manufacturing depends on operators' ability to validate and trust model predictions. Given the reliance on interpretable statistical rules and expert judgment in traditional QC, resistance to AI adoption is common. Looking ahead, as the demand for intelligent systems increases, predictive maintenance will become a key focus of QC. To meet this demand, future AI models are supposed to not only provide accurate predictions but also offer reliable and interpretable results. Ensuring high reliability and transparency will be essential for integrating AI into safety-critical and predictive maintenance applications.

Furthermore, industrial QC scenarios are highly fragmented and customized across products, lines, and factories, which complicates the notion of a unified QC foundation model. High-mix, low-volume production and frequent product changeovers further exacerbate data drift and distribution shifts between training and deployment environments, making static models brittle. In QC, this fragmentation means that LMs must be designed for efficient adaptation, e.g., through parameter-efficient fine-tuning, modular plug-ins, or line-specific adapters, rather than only relying on monolithic pre-training over a unified, large-scale data set.

Third, domain-specific challenge is the integration of rich engineering and process knowledge into LMs. Recent work proposes LLM-enabled process knowledge graphs and LLM-assisted fine-grained knowledge graph construction for manufacturing, but also makes clear that domain knowledge remains highly specialized, multi-layered, and difficult to encode in generic pre-training corpora (Xu et al., 2025). For QC specifically, relevant knowledge includes defect taxonomies, control plans, tolerances, and cause–effect chains in processes, information that is scattered across internal standards, documents, and operator experience. Bridging this gap requires hybrid architectures in which LMs are tightly coupled with manufacturing knowledge graphs, ontologies, and rule-based systems, rather than relying solely on unstructured text pre-training.

Model-level challenges hinder the deployment of LMs in industrial QC settings. To be viable on the factory floor, models must be reliable, computationally lightweight and optimized for low-latency performance. Equally critical is the injection of domain knowledge to enhance task relevance and user trust. Future LMs must balance efficiency, interpretability and contextual alignment to enable scalable, real-time and reliable quality assurance.

4.1.3 Evaluation-level

Evaluating and sustaining model performance in real-world factory environments presents a persistent and multifaceted challenge. Unlike many other AI applications, industrial QC operates in dynamic, high-stakes settings where product mixes, process conditions and inspection criteria change frequently, and where misclassifications can translate directly into scrap, rework or safety risk. Consequently, static benchmarks and periodic offline evaluations are inadequate for capturing the dynamic and heterogeneous nature of industrial operations. Instead, continuous monitoring frameworks are essential for assessing model behavior under time-varying conditions and for promptly detecting performance drift. This necessitates the development of robust and fine-grained evaluation metrics that can reflect long-term operational diversity, task complexity and defect variability.

As models are exposed to new data and shifting environments, a critical requirement for long-term deployment is enabling models to adapt to evolving operational scenarios, such as changes in product types, workflows, or inspection criteria, while preserving previously acquired knowledge. Evaluation frameworks must therefore assess both the adaptability of models to new data and their retention of prior competencies. This calls for continual learning mechanisms that balance plasticity (to incorporate new information) and stability (to prevent forgetting), thereby supporting reliable generalization across both familiar and novel contexts.

Furthermore, model evaluation should not function as an isolated post hoc analysis but must be tightly integrated into the broader QC and decision-making infrastructure. In high-throughput industrial settings, real-time model assessments should inform operational decisions, such as initiating inspections, pausing production, or tuning process parameters. This level of integration enables automated feedback loops, enhances anomaly response speed and reinforces traceability and accountability, ultimately supporting closed-loop quality assurance.

To ensure long-term model reliability and adaptability, evaluation must be positioned as a core, continuous

component of the AI deployment pipeline. Regular performance validation and adaptive recalibration are essential to mitigate performance degradation over time. However, the lack of standardized and domain-specific benchmarks that reflect the dynamic and complex nature of industrial data remains an open challenge. Addressing this gap is vital for achieving sustained model excellence in real-world manufacturing environments.

4.2 Perspectives

To fully realize the potential of LMs in industrial QC, future research must address several strategic directions aimed at enhancing data readiness, architectural design and adaptability (Fig. 4).

4.2.1 Building high-quality large-scale datasets

Addressing data challenges is a top priority for deploying foundation models in manufacturing. A primary focus lies in the construction of high-quality and large-scale industrial data sets. Unlike traditional public benchmarks, industrial data sets should be developed to capture variability across industries, production lines and defect types to support robust pretraining for generalizable model performance. To address scalability, one promising avenue is leveraging the generative and self-supervised capabilities of LMs themselves. LMs can be employed to automate data annotation, generate synthetic defect samples, or simulate rare operational conditions, thus enriching training corpora without relying solely on manual labeling. Another promising direction is the advancement of collaborative learning paradigms such as federated or distributed learning. These approaches can enable cross-organizational knowledge sharing while preserving data privacy, allowing multiple manufacturers to jointly contribute to robust and general-purpose QC models. Such collaborations can facilitate broader representation of failure modes and process diversity, accelerating model robustness and cross-domain transferability. Progress in this space will benefit from interdisciplinary collaboration across AI algorithms, manufacturing engineering and data governance. Future research may explore new standards for industrial data interoperability, improved simulation-to-reality transfer and privacy-preserving learning protocols tailored for industrial applications.

4.2.2 Integrating downstream tasks

To fully harness the transformative potential of LMs in manufacturing, future research should focus on their seamless integration into real-world production workflows. A promising direction lies in the development of modular architectures, where the LM serves as a versatile backbone, complemented by lightweight and task-specific adapters for downstream applications. This decoupled structure enhances adaptability, reduces computational overhead and facilitates rapid deployment across varied use cases. Also, enabling cross-task knowledge sharing is essential for amplifying the generalization capabilities of LMs. Insights gained from one domain can be transferred to and reinforce performance in others. This shared representational space fosters a unified intelligence layer that dynamically improves with cumulative experience.

As intelligent QC systems evolve, they must progress from simple process monitoring and fault diagnosis to more sophisticated functions like quality predicting and predictive maintenance which aims to reduce downtime and operational costs. To facilitate this transition, LMs must be enhanced to not only provide real-time diagnostic insights but also deliver reliable predictions of potential failures and quality deviations. Integrating Bayesian statistical methods and statistical modeling approaches into LMs can significantly improve the accuracy and interpretability of these predictions. By incorporating uncertainty quantification, these models can offer probabilistic forecasts, helping operators better understand the risk of failures and the likelihood of quality issues before they occur. This predictive capability is essential for adopting a proactive approach to QC.

In parallel, workflow integration remains a critical frontier. LMs must be embedded into the broader digital manufacturing ecosystem in a way that aligns with human decision-making processes, automation cycles and

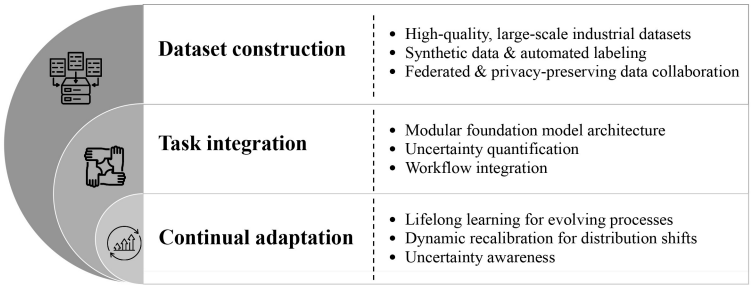

Fig. 4 Prospective research opportunities for large models in quality control.

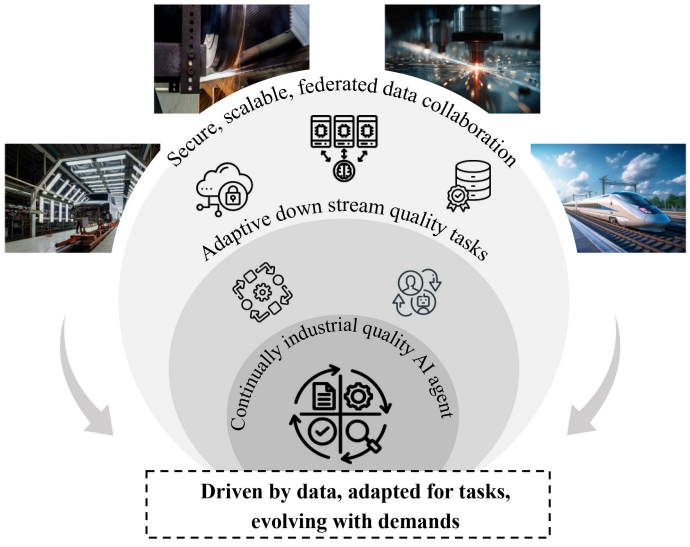

Fig. 5 Architecture of a continually adaptive industrial quality AI agent.

system-level feedback loops. The integration of LMs with DTs, for instance, can provide a dynamic representation of physical systems, improving both predictive maintenance and quality forecasting. This workflow integration requires robust human-AI interfaces, real-time model orchestration, and context-aware interpretability mechanisms to ensure that the insights provided by the models are not only accurate but also actionable.

By combining these elements, manufacturing systems can achieve a closed-loop lifecycle QC, where data flows seamlessly from production to prediction, supporting continuous improvement and ensuring the long-term reliability of manufacturing operations.

4.2.3 Continual learning and dynamic adaptation

An important direction for future research is endowing LMs with continual learning and dynamic adaptation capabilities, enabling them to evolve with the manufacturing processes. To remain effective over time, LMs must move beyond the conventional train-once-deploy-forever paradigm and adopt a lifelong learning framework that supports incremental knowledge acquisition without forgetting previously learned patterns. Future models should incorporate adaptive mechanisms that can autonomously detect distribution shifts and initiate recalibration when necessary. This will ensure sustained performance in the face of operational variability and gradual system changes. Collectively, these capabilities will drive the development of resilient and self-improving AI systems that are capable of continuous alignment with real-world manufacturing processes, paving the way for more intelligent, autonomous and trustworthy industrial operations.

As illustrated in Fig. 5, an effective path forward lies in developing a continually evolving industrial AI agent fueled by secure, scalable data collaboration, capable of adapting across diverse downstream tasks and sustained by lifelong learning. This integrated framework offers a foundation for building resilient, generalizable and self-improving AI systems aligned with the dynamic nature of industrial environments.

5 Conclusions

LMs offer transformative potential for the QC domain, empowering manufacturing systems to become more intelligent and autonomous. Their strengths in multimodal perception, structured reasoning and human-AI collaboration can substantially elevate quality-centric processes. Moreover, real-world challenges in the manufacturing domain offer LMs concrete and application-oriented problem scenarios that both challenge and enrich their development. Fully realizing the potential demands progress across model design, deployment and lifecycle management, with future research focusing on transparency, adaptability and seamless industrial integration to ensure sustained impact.

Competing Interests The authors declare that they have no competing interests.

References

Abhilash P M, Luo X, Liu Q, Madarkar R, Walker C (2024). Towards next-gen smart manufacturing systems: The explainability revolution. npj. Advances in Manufacturing, 1(1): 8

Ahmed S, Nielsen I E, Tripathi A, Siddiqui S, Ramachandran R P, Rasool G (2023). Transformers in time-series analysis: A tutorial. Circuits, Systems, and Signal Processing, 42(12): 7433–7466

Alayrac J B, Donahue J, Luc P, Miech A, Barr I, Hasson Y, Lenc K, Mensch A, Millican K et al (2022). Flamingo: a visual language model for few-shot learning. In: Proceedings of Advances in Neural Information Processing Systems. New Orleans, USA: Curran Associates, Inc., 35: 23716–23736

Atrey P K, Hossain M A, El Saddik A, Kankanhalli M S (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Systems, 16(6): 345–379

Badini S, Regondi S, Frontoni E, Pugliese R (2023). Assessing the capabilities of ChatGPT to improve additive manufacturing troubleshooting. Advanced Industrial and Engineering Polymer Research, 6(3): 278–287

Baltrušaitis T, Ahuja C, Morency L P (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2): 423–443

Bergmann P, Jin X, Sattlegger D, Steger C (2021). The MVTec 3D-AD dataset for unsupervised 3D anomaly detection and localization. Preprint at arXiv. arXiv:2112.09045

Bi X, Chen D, Chen G, Chen S, Dai D, Deng C, Ding H, Dong K, Du Q, Fu Z et al (2024). DeepSeek LLM: Scaling open-source language models with longtermism. Preprint at arXiv. arXiv:2107.03374

Bianchini F, Calamo M, Marinacci M, Rossi J, Mecella M (2025). Automating industrial quality control: a multimodal LLM and RAG framework for anomaly detection. In: Artificial Intelligence Appiclations and Innovations April. Cham: Springer, 758: 253–266

Bommasani R, Hudson D A, Adeli E, Altman R, Arora S, von Arx S, Bernstein M S, Bohg J, Bosselut A et al (2022). On the opportunities and risks of foundation models. Preprint at arXiv. arXiv:2108.07258

Brown T B, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry et al (2020). Language models are few-shot learners. In: Proceedings of Advances in Neural Information Processing Systems December. Vancouver, Canada: Curran Associates, Inc., 33: 1877–1901

Cao K, Zhang T, Huang J (2024). Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems. Scientific Reports, 14(1): 4890

Cao Y, Xu X, Cheng Y, Sun C, Du Z, Gao L, Shen W (2025). Personalizing vision-language models with hybrid prompts for zero-shot anomaly detection. IEEE Transactions on Cybernetics, 55(4): 1917–1929

Capezza C, Capizzi G, Centofanti F, Lepore A, Palumbo B (2025). An adaptive multivariate functional EWMA control chart. Journal of Quality Technology, 57(1): 1–15

Capezza C, Centofanti F, Lepore A, Palumbo B (2024). Robust multivariate functional control chart. Technometrics, 66(4): 531–547

Caruana R (1997). Multitask learning. Machine Learning, 28(1): 41–75

Chen X, Djolonga J, Padlewski P, Mustafa B, Changpinyo S, Wu J, Ruiz C R, Goodman S, Wang X (2024a). On scaling up a multilingual vision and language model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 14432–14444

Chen X, Shi S, Ma T, Zhou J, See S, Cheung K C, Li H (2025). M3net: Multimodal multi-task learning for 3D detection, segmentation, and occupancy prediction in autonomous driving. In: Proceedings of AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI Press, 39: 2275–2283

Chen M, Tworek J, Jun H, Yuan Q, de Oliveira Pinto H P, Kaplan J, Edwards H, Burda Y, Joseph N et al (2021). Evaluating large language models trained on code. Preprint at arXiv. arXiv:2107.03374

Chen S, Zhang Y, Yang Q (2024b). Multi-task learning in natural language processing: An overview. ACM Computing Surveys, 56(12): 1–32

Chkirbene Z, Hamila R, Gouissem A, Devrim U (2024). Large Language Models in Industry: A survey of applications, challenges, and trends. In: Proceedings of the 21st IEEE International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT. Charlotte, USA: IEEE, 229–234

Costanzino A, Ramirez P Z, Lisanti G, Di Stefano L (2024). Multimodal industrial anomaly detection by crossmodal feature mapping. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 17234–17243

Deng J, Liu G, Wang L, Yuan B, Huang H (2023). Research progress of intelligent optimization design of manufacturing process parameters. Manufacturing Technology & Machine Tool, 0(5): 74–80

Devlin J, Chang M W, Lee K, Toutanova K (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA: Association for Computational Linguistics, 1: 4171–4186

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2020). An image is worth 16×16 words: Transformers for image recognition at scale. Preprint at arXiv. arXiv:2010.11929

Ebadi M, Chenouri S, Lin D K J, H. Steiner S. (2021). Statistical monitoring of the covariance matrix in multivariate processes: A literature review. Journal of Quality Technology, 54(3): 269–289

El-Sappagh S, Abuhmed T, Riazul Islam S M, Kwak K S (2020). Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. Neurocomputing, 412(28): 197–215

Eldele E, Ragab M, Qing X Edward, Chen Z, Wu M, Li X, Lee J (2025). UniFault: A fault diagnosis foundation model from bearing data. Preprint at arXiv. arXiv:2504.01373

Era I Z, Ahmed I, Liu Z, Das S (2024). An unsupervised approach towards promptable defect segmentation in laser-based additive manufacturing by segment anything. Preprint at arXiv. arXiv:2312.04063

Escobar C A, Cantoral-Ceballos J A, Morales-Menendez R (2025). Quality 4.0: Learning quality control, the evolution of SQC/SPC. Quality Engineering, 37(1): 92–117

Fang Q, Xiong G, Wang F, Shen Z, Dong X, Wang F (2024). Large language models as few-shot defect detectors for additive manufacturing. In: Proceedings of China Automation Congress . Qingdao, IEEE: 6900–6905

Fatima S S W, Rahimi A (2024). A review of time-series forecasting algorithms for industrial manufacturing systems. Machines, 12(6): 380–400

Gaikwad A, Yavari R, Montazeri M, Cole K, Bian L, Rao P (2020). Toward the digital twin of additive manufacturing: Integrating ther-

mal simulations, sensing, and analytics to detect process faults. IISE Transactions, 52(11): 1204–1217

Gao S, Koker T, Queen O, Hartvigsen T, Tsiligkaridis T, Zitnik M (2024). UniTS: A unified multi-task time series model. In: Proceedings of Advances in Neural Information Processing Systems. Vancouver, Canada: Curran Associates, Inc., 37: 140589–140631

Gomaa A H (2025). RCM 4.0: A novel digital framework for reliability-centered maintenance in smart industrial systems. International Journal of Emerging Science and Engineering, 13(5): 32–43

Gu Z, Zhu B, Zhu G, Chen Y, Tang M, Wang J (2024). AnomalyGPT: Detecting industrial anomalies using large vision-language models. In: Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Press, 38(3): 1932–1940

Guo J, Yan H, Zhang C (2023). A bayesian partially observable online change detection approach with Thompson sampling. Technometrics, 65(2): 179–191

Guo M, Xu T, Liu J, Liu Z, Jiang P, Mu T, Zhang S H, Martin R R, Cheng M M, Hu S M (2022). Attention mechanisms in computer vision: A survey. Computational Visual Media, 8(3): 331–368

Han Y, Zhang F, Li Z, Wang Q, Li C, Lai P, Li T, Teng F, Jin Z (2025). MT-ConvFormer: A multitask bearing fault diagnosis method using a combination of CNN and Transformer. IEEE Transactions on Instrumentation and Measurement, 74: 3501816

Ho S L, Xie M (1998). The use of ARIMA models for reliability forecasting and analysis. Computers & Industrial Engineering, 35(1–2): 213–216

Hu T, Zhang J, Yi R, Du Y, Chen X, Liu L, Wang Y, Wang C (2024). Anomaly diffusion: Few-shot anomaly image generation with diffusion model. In: Proceedings of AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press, 38: 8526–8534

Hu Y, Miao X, Si Y, Pan E, Zio E (2022). Prognostics and health management: A review from the perspectives of design, development and decision. Reliability Engineering & System Safety, 217: 108063

Huang Y, Wang E H, Liu Z, Pan L, Li H, Liu X (2023). Modeling task relationships in multivariate soft sensor with balanced mixture-of-experts. IEEE Transactions on Industrial Informatics, 19(5): 6556–6564

Huang Y, Zhu J, Zhong X, Deng Y (2025). SAID: segment all industrial defects with scene prompts. Sensors, 25(16): 4929

Jacobs R A, Jordan M I, Nowlan S J, Hinton G E (1991). Adaptive mixtures of local experts. Neural Computation, 3(1): 79–87

Jeong J, Zou Y, Kim T, Zhang D, Ravichandran A, Dabeer O (2023). Winclip: Zero-/few-shot anomaly classification and segmentation. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 19606–19616

Jia C, Yang Y, Xia Y, Chen Y, Parekh Z, Pham H, Le Q, Sung Y, Li Z, Duerig T (2021). Scaling up visual and vision-language representation learning with noisy text supervision. In: Proceedings of the 38th International Conference on Machine Learning. PMLR, 139: 4904–4916

Jiang F, Qin C, Yao K, Fang C, Zhuang F, Zhu H, Xiong H (2024). Enhancing question answering for enterprise knowledge bases using large language models. In: Proceedings of International Conference on Database Systems for Advanced Applications April. Singapore: Springer, 13941: 273–290

Jiao T, Guo C, Feng X, Chen Y, Song J (2024). A comprehensive survey on deep learning multi-modal fusion: methods, technologies and applications. Computers, Materials & Continua, 80(1): 1–35

Jignasu A, Marshall K, Ganapathysubramanian B, Balu A, Hegde C, Krishnamurthy A (2023). Towards foundational AI models for additive manufacturing: language models for G-code debugging, manipulation, and comprehension. Preprint at arXiv. arXiv:2309.02465

Jin Q, Jiang Y, Lu X, Liu Y, Chen Y, Gao D, Sun Q, Zhuo C (2024a). SEM-CLIP: Precise few-shot learning for nanoscale defect detection in scanning electron microscope image. In: Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design. New Jersey: IEEE/ACM, 134: 1–8

Jin M, Wang S, Ma L, Chu Z, Zhang J, Shi X, Chen P, Liang Y, Li Y, Pan S, Wen Q (2024b). Time-LLM: Time series forecasting by reprogramming large language models. In: International Conference on Representation Learning. Vienna: OpenReview, 23857–23880

Kawaharazuka K, Obinata Y, Kanazawa N, Okada K, Inaba M (2023). Robotic applications of pre-trained vision-language models to various recognition behaviors. In: Proceedings of IEEE-RAS 22nd International Conference on Humanoid Robots. Austin: IEEE, 22: 1–8

Kim D, Park J, Lee J, Kim H (2024). Are self-attentions effective for time series forecasting? In: Proceedings of Advances in Neural Information Processing Systems December. Vancouver: Curran Associates, Inc., 37: 114180–114209

Kim G, Lim H, Kim Y, Kwon O, Choi J H (2023). Intra-person multitask learning method for chronic-disease prediction. Scientific Reports, 13(1): 1069

Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg A, Lo W, Dollar P, Girshick R (2023). Segment anything. In: Proceedings of IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE, 4015–4026

Klingenberg C O, Borges M A V, Antunes J A V Jr (2019). Industry 4.0 as a data-driven paradigm: A systematic literature review on technologies. Journal of Manufacturing Technology Management, 32(3): 570–592

Kottapalli S R K, Hubli K, Chandrashekhara S, Jain G, Hubli S, Botla G, Doddaiah R (2025). Foundation models for time series: A survey. Preprint at arXiv. arXiv:2504.04011

Kouchakzadeh A, ElMaraghy W (2024). The effect of fault detection, diagnosis, and recovery on resilience in manufacturing systems. International Journal of Advanced Manufacturing Technology, 135(11–12): 5893–5909

Kernan Freire S, Wang C, Foosherian M, Wellsandt S, Ruiz-Arenas S, Niforatos E (2024). Knowledge sharing in manufacturing using LLM-powered tools: User study and model benchmarking. Frontiers in Artificial Intelligence, 7: 1293084

Kudo T, Richardson J (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Brussels: Association for Computational Linguistics, 66–71

Lang Q, Tian S, Wang M, Wang J (2024). Exploring the answering capability of large language models in addressing complex knowledge in entrepreneurship education. IEEE Transactions on Learning Technologies, 17: 2053–2062

Lee S M, Lee D, Kim Y S (2019). The quality management ecosystem

for predictive maintenance in the industry 4.0 era. International Journal of Quality Innovation, 5(1): 4

Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi A K (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing, 138: 106587

Leng J, Li R, Xie J, Zhou X, Li X, Liu Q, Chen X, Shen W, Wang L (2025). Federated learning-empowered smart manufacturing and product lifecycle management: A review. Advanced Engineering Informatics, 65: (Part A)103179

Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W T, Rocktäschel T, Riedel S, Kiela D (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proceedings of Advances in Neural Information Processing Systems. Virtual Event: Curran Associates, Inc., 33: 9459–9474

Li Y, Du J, Jiang W (2024a). Reinforcement learning for process control with application in semiconductor manufacturing. IISE Transactions, 56(6): 585–599

Li S, Tang H (2024). Multimodal alignment and fusion: A survey. Preprint at arXiv. arXiv:2411.17040

Li Y F, Wang H, Sun M (2024b). ChatGPT-like large-scale foundation models for prognostics and health management: A survey and roadmaps. Reliability Engineering & System Safety, 243: 109850

Li J, Xie Y, Tian Y, Yin Z, Sun Z, Zhang W (2024c). Industrial process fault diagnosis based on video recognition and multi-source information fusion. Chemical Engineering Research & Design, 208: 820–836

Li Y, Zhao H, Jiang H, Pan Y, Liu Z, Wu Z, Shu P, Tian J, Yang T, Xu S et al. (2024d). Large language models for manufacturing. Preprint at arXiv. arXiv:2410.21418

Li H, Zhang Q, Li W, Liang X (2024e). Multi-modal quality prediction algorithm based on anomalous energy tracking attention. In: Proceedings of International Conference on Intelligent Computing. Fuzhou, China: Springer, 150–162

Li Q, Zhang X, Huang J, He H, Zhang F, Qin Z, Chu F (2024f). VSLLaVA: A pipeline of large multimodal foundation model for industrial vibration signal analysis. Preprint at arXiv. arXiv:2409.07482

Liang X, Zhang M, Feng G, Wang D, Xu Y, Gu F (2023). Few-shot learning approaches for fault diagnosis using vibration data: a comprehensive review. Sustainability, 15(20): 14975

Lim J, Vogel-Heuser B, Kovalenko I (2024). Large language model-enabled multi-agent manufacturing systems. In: Proceedings of IEEE 20th International Conference on Automation Science and Engineering August. Lyon, France: IEEE, 3940–3946

Liu S, Bao J, Zheng P (2023a). A review of digital twin-driven machining: From digitization to intellectualization. Journal of Manufacturing Systems, 67: 361–378

Liu X, He P, Chen W, Gao J (2019). Multi-task deep neural networks for natural language understanding. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 57: 4487–4496.

Liu F, Li G, Zhao Y, Jin Z (2020). Multi-task learning based pre-trained language model for code completion. In: Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering. Melbourne, Australia: IEEE, 473–485

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of IEEE/CVF International Conference on Computer Vision October. Montreal, Canada: IEEE, 10012–10022

Liu D, Wang Y, Liu C, Yuan X, Yang C, Gui W (2023b). Data mode related interpretable transformer network for predictive modeling and key sample analysis in industrial processes. IEEE Transactions on Industrial Informatics, 19(9): 9325–9336

Liu D, Wang Y, Liu C, Yuan X, Yang C (2024). Multirate-Former: An efficient transformer-based hierarchical network for multistep prediction of multirate industrial processes. IEEE Transactions on Instrumentation and Measurement, 73: 2502313

Lu J, Batra D, Parikh D, Lee S (2019). ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In: Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver: Curran Associates, Inc., 2: 11

Lukens S, McCabe L H, Gen J, Ali A (2024). Large Language Model Agents as Prognostics and Health Management Copilots. Annual Conference of the PHM Society, 16: (1)

Lv Y, Zhang X, Cheng Y, Lee C (2024). Intelligent fault diagnosis of machinery based on hybrid deep learning with multi temporal correlation feature fusion. Quality and Reliability Engineering International, 40(6): 3517–3536

Madaan A, Tandon N, Gupta P, Hallinan S, Gao L, Wiegreffe S, Alon U, Dziri N, Prabhumoye Set al (2023). Self-refine: Iterative refinement with self-feedback. In: Proceedings of Advances in Neural Information Processing Systems. New Orleans, USA: Curran Associates, Inc., 46534–46594

Mahesh N, Devishamani C S, Raghu K, Mahalingam M, Bysani P, Chakravarthy A V, Raman R (2024). Advancing healthcare: the role and impact of AI and foundation models. American Journal of Translational Research, 16(6): 2166–2179

Masserano L, Ansari A F, Han B, Zhang X, Faloutsos C, Mahoney M W, Wilson A G, Park Y, Rangapuram S, Maddix D C, Wang Y (2024). Enhancing foundation models for time series forecasting via wavelet-based tokenization. In: Proceedings of the 42nd International Conference on Machine Learning. Vancouver: PMLR, 267: 43248–43275

McKinney M, Garland A, Cillessen D, Adamczyk J, Bolintineanu D, Heiden M, Fowler E, Boyce B L (2025). Unsupervised multimodal fusion of in-process sensor data for advanced manufacturing process monitoring. Journal of Manufacturing Systems, 78: 271–282

Megahed F M, Chen Y J, Colosimo B M, Grasso M L G, Jones-Farmer L A, Knoth S, Sun H, Zwetsloot I (2025). Adapting OpenAI's CLIP model for few-shot image inspection in manufacturing quality control: An expository case study with multiple application examples. Preprint at arXiv. arXiv:2501.12596

Megahed F M, Chen Y J, Zwetsloot I M, Knoth S, Montgomery D C, Jones-Farmer L A (2024). Introducing ChatSQC: Enhancing statistical quality control with augmented AI. Journal of Quality Technology, 56(5): 474–497

Moosavi S, Farajzadeh-Zanjani M, Razavi-Far R, Palade V, Saif M (2024). Explainable AI in manufacturing and industrial cyber–physical systems: A survey. Electronics, 13(17): 3497

Nagrani A, Yang S, Arnab A, Jansen A, Schmid C, Sun C (2021). Attention bottlenecks for multimodal fusion. In: Advances in

Neural Information Processing Systems. Curran Associates, Inc., 34: 14200–14213

O'Leary D E (2023). Enterprise large language models: Knowledge characteristics, risks, and organizational activities. Intelligent Systems in Accounting, Finance & Management, 30(3): 113–119

Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C L, Mishkin P, Zhang C, Agarwal S, Slama K et al (2022). Training language models to follow instructions with human feedback. In: Proceedings of Advances in Neural Information Processing Systems. New Orleans, USA: Curran Associates, Inc., 35: 27730–27744

Peng W, Li G, Jiang Y, Wang Z, Ou D, Zeng X, Chen E (2024). Large language model based long-tail query rewriting in Taobao search. In: Proceedings of Companion Proceedings of the ACM Web Conference 2024. Singapore: ACM, 20–28

Peršak E, Anjos M F, Lautz S, Kolev A (2024). Multiple-resolution tokenization for time series forecasting with an application to pricing. Preprint at arXiv. arXiv:2407.03185

Psarommatis F, May G (2023). A literature review and design methodology for digital twins in the era of zero-defect manufacturing. International Journal of Production Research, 61(16): 5723–5743

Qiu P, Xie X (2022). Transparent sequential learning for statistical process control of serially correlated data. Technometrics, 64(4): 487–501

Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021). Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning. Virtual Event: 8748–8763

Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu P J (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140): 1–67

Rasul K, Ashok A, Williams AR, Ghonia H, Bhagwatkar R, Khorasani A, Bayazi MJ, Adamopoulos G, Riachi R, Hassen N, Biloš M (2023). Lag-llama: Towards foundation models for probabilistic time series forecasting. Preprint at arXiv. arXiv:2310.08278

Sahoo P, Meharia P, Ghosh A, Saha S, Jain V, Chadha A (2024). A comprehensive survey of hallucination in large language, image, video and audio foundation models. In: Findings of the Association for Computational Linguistics: EMNLP November. Miami: Association for Computational Linguistics, 11709–11724

Schuster M, Nakajima K (2012). Japanese and Korean voice search. In: Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing March. Kyoto: IEEE, 5149–5152

Sennrich R, Haddow B, Birch A (2016). Neural machine translation of rare words with subword units. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics August. Berlin: Association for Computational Linguistics, 1715–1725

Singhal K, Azizi S, Tu T, Mahdavi S S, Wei J, Chung H W, Scales N, Tanwani A, Cole-Lewis H et al. (2023). Large language models encode clinical knowledge. Nature, 620(7972): 172–180

Singhal K, Tu T, Gottweis J, Sayres R, Wulczyn E, Amin M, Hou L, Clark K, Pfohl S Ret al., (2025). Toward expert-level medical question answering with large language models. Nature Medicine, 31(3): 943–950

Song K, Cui W, Yu H, Li X, Yan Y (2024). SAM Era: Can it segment any industrial surface defects? Computers, Materials & Continua, 78: (3)3953–3969

Su J, Jiang C, Jin X, Qiao Y, Xiao T, Ma H, Wei R, Jing Z, Xu J, Lin J (2024). Large language models for forecasting and anomaly detection: A systematic literature review. Preprint at arXiv. arXiv:2402.10350

Sun J, Liao Q V, Muller M, Agarwal M, Houde S, Talamadupula K, Weisz D (2022). Investigating explainability of generative AI for code through scenario-based design. In: Proceedings of the 27th International Conference on Intelligent User Interfaces March. Helsinki, Finland: ACM, 212–228

Talukder S, Yue Y, Gkioxari G (2024). TOTEM: Tokenized time series embeddings for general time series analysis. Preprint at arXiv. arXiv:2402.16412

Tan H, Bansal M (2019). LXMERT: Learning cross-modality encoder representations from transformers. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing November. Hong Kong: Association for Computational Linguistics, 5100–5111

Tercan H, Meisen T (2022). Machine learning and deep learning based predictive quality in manufacturing: A systematic review. Journal of Intelligent Manufacturing, 33(7): 1879–1905

Tsai Y H, Bai S, Liang P P, Kolter J Z, Morency L, Salakhutdinov R (2019). Multimodal transformer for unaligned multimodal language sequences. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 6558–6569

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, Polosukhin I (2017). Attention is all you need. In: Proceedings of Advances in Neural Information Processing Systems December. California: Curran Associates, 4–9: 5998–6008

van Dinter R, Tekinerdogan B, Catal C (2022). Predictive maintenance using digital twins: A systematic literature review. Information and Software Technology, 151: 107008

Wang C, Zhu H, Peng J, Wang Y, Yi R, Wu Y, Ma L, Zhang J (2025). M3DM-NR: RGB-3D noisy-resistant industrial anomaly detection via multimodal denoising. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(11): 9981–9993

Wang H, Li Y, Xie M (2024a). Empowering ChatGPT-like large-scale language models with local knowledge base for industrial prognostics and health management. Preprint at arXiv. arXiv:2312.14945

Wang P, Qu H, Zhang Q, Xu X, Yang S (2023a). Production quality prediction of multistage manufacturing systems using multi-task joint deep learning. Journal of Manufacturing Systems, 70: 48–68

Wang X, Chen G, Qian G, Gao P, Wei X Y, Wang Y, Tian Y, Gao W. (2023b). Large-scale multi-modal pre-trained models: A comprehensive survey. Machine Intelligence Research, 20(4): 447–482

Wang X, Zhang X, Cao Y, Wang W, Shen C, Huang T (2023c). SegGPT: Segmenting everything in context. Preprint at arXiv. arXiv:2304.03284

Wang Y, Dai R, Liu D, Wang K, Yuan X, Liu C (2024b). A task-oriented deep learning framework based on target-related transformer network for industrial quality prediction applications. Engineering

Applications of Artificial Intelligence, 133: (Part D)108361

Wen Q, Zhou T, Zhang C, Chen W, Ma Z, Yan J, Sun L (2022). Transformers in time series: A survey. In: Proceedings of the 39th International Conference on Machine Learning July. Baltimore, USA: ACM, 39: 1–15

Woodall W H, Montgomery D C (2014). Some current directions in the theory and application of statistical process monitoring. Journal of Quality Technology, 46(1): 78–94

Wu W, Peng W, Liu J, Li X, Zhang D, Sun J (2025a). An attention-based weight adaptive multi-task learning framework for slab head shape prediction and optimization during the rough rolling process. Journal of Manufacturing Processes, 133(17): 408–429

Wu G, Zhang Y, Deng L, Zhang J, Chai T (2025b). Cross-modal learning for anomaly detection in complex industrial processes: Methodology and benchmark. IEEE Transactions on Circuits and Systems for Video Technology, 35(3): 2632–2645

Wu L, Zheng Z, Qiu Z, Wang H, Gu H, Shen T, Qin C, Zhu C, Zhu H, Liu Q, Xiong H, Chen E (2024). A survey on large language models for recommendation. World Wide Web (Bussum), 27(5): 60

Wu Y, Meng Y, Shao C (2022). End-to-end online quality prediction for ultrasonic metal welding using sensor fusion and deep learning. Journal of Manufacturing Processes, 83: 685–694

Xie Z, Chen J, Feng Y, Zhang K, Zhou Z (2022). End to end multi-task learning with attention for multi-objective fault diagnosis under small sample. Journal of Manufacturing Systems, 62: 301–316

Xu D, Chen W, Peng W, Zhang C, Xu T, Zhao X, Wu X, Zheng Y, Wang Y, Chen E (2024). Large language models for generative information extraction: A survey. Frontiers of Computer Science, 18(6): 186357

Xu Q, Qiu F, Zhou G, Zhang C, Ding K, Chang F, Lu F, Yu Y, Ma D, Liu J (2025). A large language model-enabled machining process knowledge graph construction method for intelligent process planning. Adv Eng Inform, 65: (Part B)103244

Yan H, Sergin N D, Brenneman W A, Lange S J, Ba S (2021). Deep multistage multi-task learning for quality prediction of multistage manufacturing systems. Journal of Quality Technology, 53(5): 526–544

Yang T, Chang L, Yan J, Li J, Wang Z, Zhang K (2025). A survey on foundation-model-based industrial defect detection. Preprint at arXiv. arXiv:2502.19106

Yang X, Zhang C (2024). Online directed-structural change-point detection: A segment-wise time-varying dynamic Bayesian network approach. IISE Transactions, 56(5): 527–540

Yang Y, Gao R, Tang Y, Antic S L, Deppen S, Huo Y, Sandler K, Massion P, Landman B (2020). Internal-transfer weighting of multi-task learning for lung cancer detection. In: Proceedings of SPIE International Society for Optics and Photonics. Houston, USA: SPIE, 10

Yin S, Ding S X, Xie X, Luo H (2014). A review on basic data-driven approaches for industrial process monitoring. IEEE Transactions on Industrial Electronics, 61(11): 6418–6428

Yu Y, Xue J, Dai S, Bao Q, Zhao G (2021). Quality prediction and process parameter optimization method for machining parts. Journal of Zhejiang University (Engineering Science), 55(3): 44–51

Yuan X, Lin Z, Kuen J, Zhang J, Wang Y, Maire M, Kale A, Faieta B (2021). Multimodal contrastive training for visual representation learning. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition June. Nashville, USA: IEEE, 6995–7004

Zajec P, Rožanec J M, Theodoropoulos S, Fontul M, Koehorst E, Fortuna B, Mladenić D (2024). Few-shot learning for defect detection in manufacturing. International Journal of Production Research, 62(19): 6979–6998

Zhai X, Wang X, Mustafa B, Steiner A, Keysers D, Kolesnikov A, Beyer L (2022). LiT: Zero-shot transfer with locked-image text tuning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition June. New Orleans: IEEE, 18102–18112

Zhang H, Dereck S S, Wang Z, Lv X, Xu K, Wu L, Jia Y, Wu J, Long Z, Liang W, Ma X G, Zhuang R (2026). Large scale foundation models for intelligent manufacturing applications: A survey. J Intell Manuf, 37: 119–170

Zhang Y, Yang Q (2022). A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 34(12): 5586–5609

Zhao F, Zhang C, Geng B (2024). Deep multimodal data fusion. ACM Computing Surveys, 56(9): 1–36

Zhong R, Lee K, Zhang Z, Klein D (2021). Adapting language models for zero-shot learning by meta-tuning on dataset and prompt collections. In: Findings of the Association for Computational Linguistics: EMNLP 2021 November. Punta Cana: Association for Computational Linguistics, 2856–2878

Zhou B, Li X, Liu T, Xu K, Liu W, Bao J (2024). CausalKGPT: Industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing. Advanced Engineering Informatics, 59: 102333

Zhou L, Wang H (2024). An adaptive multi-scale feature fusion and adaptive mixture-of-experts multi-task model for industrial equipment health status assessment and remaining useful life prediction. Reliability Engineering & System Safety, 248: 110190

Zhu X, Zhang R, He B, Guo Z, Zeng Z, Qin Z, Zhang S, Gao P (2023). PointCLIP v2: Prompting CLIP and GPT for powerful 3D open-world learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision October. Paris: IEEE, 2023: 2639–2650.

Zuo Z, Dong J, Wu Y, Qu Y, Wu Z (2024). Clip3D-AD: Extending CLIP for 3D few-shot anomaly detection with multi-view images generation. Preprint at arXiv. arXiv:2406.18941