Journal of Manufacturing Systems 86 (2026) 913–941

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Review

Human-in-the-loop and large language models in smart manufacturing: Current applications, challenges, and perspectives

Mahdi Sadeqi Bajestani [a], Duhwan Mun [b,*], Duck Bong Kim [c,*]

[a] UHasselt, School of Transportation Sciences, Transportation Research Institute (IMOB), Martelarenlaan 42, Hasselt 3500, Belgium
[b] School of Mechanical Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea
[c] School of Environmental, Civil, Agricultural, and Mechanical Engineering, University of Georgia, Athens, GA 30602, USA

ARTICLE INFO

Keywords:
Large Language Models
Smart Manufacturing
Human-in-the-Loop
Generative Artificial Intelligence, Digital Twin,
Cyber-Physical System

ABSTRACT

The advent of Large Language Models (LLMs) has led to a transformative shift in modern industrial production, enabling enhanced automation, decision-making, and adaptability. LLMs have demonstrated their potential in diverse manufacturing applications, including design for manufacturability, process planning, and anomaly detection. This paper presents a systematic scoping review of large language model applications in smart manufacturing with a particular emphasis on human-in-the-loop concepts. By analyzing recent industrial and academic studies, the review examines how LLMs are integrated across manufacturing functions, how human roles are defined and operationalized, and what limitations remain in terms of trust, validation, and human centricity. Based on the findings of the review, an exemplar conceptual structure is synthesized to organize existing approaches and highlight research gaps from an Industry 5.0 perspective. This four-modular structure integrates human-in-the-loop (HITL), cyber-physical systems (CPS), LLM, and verification, validation, and uncertainty management (VV&UM). The HITL module addresses real-time human oversight, refining AI-generated insights and reducing decision-making errors; the CPS module aims to bridge the gap between digital twins, real-world sensor data, and AI-driven predictions, enabling real-time validation of AI recommendations; the LLM module is responsible to enhances manufacturability awareness; and the VV&UM module establishes structured verification approach, uncertainty quantification, risk assessment mechanisms, and authentication and authorization mechanisms to ensure the AI-generated outputs are reliable and compliant with industry standards. By integrating these four modules, the LLM with human-in-the-loop-based smart manufacturing (LLM-HSM) exemplar conceptual structure creates a hybrid intelligence model where AI enhances automation, while human expertise ensures contextual accuracy, performance assurance, and quality control. This paper explores the potentials, challenges, and future perspectives of LLMs and HITL in smart manufacturing, outlining a forward-looking exemplar conceptual structure for their responsible and practical implementation in next-generation industrial systems.

1. Introduction

Smart Manufacturing (SM) uses interconnected machines and tools to enhance performance, optimize energy use, and support workforce efficiency [1]. It aims to reduce operational inefficiencies and maximize overall equipment effectiveness by integrating advanced automation and intelligent decision-making systems [2,3]. While fully automated processes can streamline shop-floor operations and mitigate challenges caused by skill shortages, they also face limits in adaptability, real-time decision-making, and human intuition. Studies demonstrate that automation technologies often displace repetitive or low-skilled tasks but simultaneously elevate demand for skilled human judgement, creative problem-solving, and oversight roles [4]. In process industries, maintenance costs remain exceedingly high due to job interruption, failures, and corrective actions, all of which often necessitate human diagnosis and intervention, and they account for substantial losses despite high levels of automation [5]. Furthermore, surveys of smart factory implementations reveal that despite increasing robot adoption, human operators continue to play central roles in supervisory, decision support, and exception-handling tasks [6].

Therefore, limitations of full automation are not merely theoretical but measurable, underscoring the need for approaches that augment

* Corresponding authors.
E-mail addresses: dhmun@korea.ac.kr (D. Mun), DBKim@uga.edu (D.B. Kim).

https://doi.org/10.1016/j.jmsy.2026.04.027
Received 5 July 2025; Received in revised form 10 March 2026; Accepted 9 April 2026
Available online 15 April 2026

rather than replace human expertise. Against this backdrop, Industry 5.0 extends the principles of Industry 4.0 by moving beyond efficiency and automation to emphasize broader human- and society-oriented priorities. It is defined as a paradigm that complements technological innovation with human-centricity, sustainability, and resilience [7]. This includes not only workplace safety, mental well-being, and environmentally sustainable practices but also resilience in supply chains, personalization of products and services, and value-driven innovation [8]. In this vision, advanced technologies such as AI, robotics, and digital twins serve as enablers that augment rather than replace human creativity, adaptability, and decision-making. By situating humans at the center of industrial ecosystems, Industry 5.0 promotes collaboration between people and intelligent systems to create more flexible, responsible, and sustainable manufacturing processes [9]. This shift is reflected in contemporary frameworks such as Industry 5.0, Society 5.0, and Operator 5.0, all of which blur the traditional boundaries between humans and machines.

To better achieve the Industry 5.0 principles, the human-in-the-loop (HITL) concept, a vital characteristic of Industry 5.0, has emerged as a paradigm that emphasizes the collaborative synergy between human operators and intelligent systems in smart manufacturing. It aims to ensure a balance between machine precision and human intuition in the decision-making processes [10]. It attempts to incorporate humans' strengths, such as human brainpower [11,12], creativity, and fault tolerance abilities, into the manufacturing process. It promotes flexibility, agility, and robustness by integrating human abilities and system capabilities. Additionally, it prioritizes core human needs and interests within the industry. This approach emphasizes not only the protection of physical health and safety in the workplace but also the importance of mental health and overall well-being. It is essential to ensure that the deployment of technological advancements does not compromise human rights at any stage [13].This includes safeguarding personal and industrial data, protecting privacy in human-AI interactions, and preventing bias or discrimination in decision-making processes, which leads us toward the realization of trustworthy AI [14] and responsible AI [15].

In smart manufacturing, humans provide irreplaceable and unique values that automation cannot replicate, such as creativity, dexterity, cognition, and decision-making abilities [16,17]. Industry 5.0 emphasizes an HITL approach to enhance the well-being and needs of human operators. This requires manufacturing processes that incorporate advanced natural interaction capabilities along with improved perception, cognition, and action intelligence. As a result, human workers can form a versatile partnership with machines and technologies, leading to higher productivity than what can be achieved through purely manual labor or automation alone. This collaborative approach can be applied across various manufacturing tasks, including inspections, diagnostics, assembly, maintenance, and logistics [18–20]. However, a straightforward interaction approach, defined here as a method that allows users of all levels of expertise, as well as those with varying physical or cognitive abilities, i.e., disabled, capability-lessening, or normal [21], to communicate with machines in an intuitive, accessible manner without requiring advanced technical training, is still in its early stages. Existing methods, such as graphical user interfaces [22], menu-driven programming [23], or pre-defined command libraries [24], often offer limited accessibility and require significant prior knowledge or training.

To address this, integrating advanced technologies, such as generative artificial intelligence (GenAI) and, more specifically, large language models (LLMs), has become paramount to achieving greater efficiency, flexibility, and innovation. LLMs offer unique capabilities in natural language understanding and multimodal data processing [25]. They aim to enhance human-computer interaction by enabling machines to understand, generate, and respond to natural language in a way that mimics human reasoning, providing valuable insights, automation, and improved decision-making across various domains [26]. LLMs predict and generate text by learning patterns from extensive textual data, enabling them to perform tasks such as text completion, translation, and summarization [27].

In this paper, based on the literature, the authors have made an effort to synthesize an exemplar conceptual structure, hereafter referred to as large language model integrated with human-in-the-loop-based smart manufacturing (LLM-HSM), to emphasize the central role of human decision-making and adaptability in cyber-physical production environments. This is achieved through integration with LLMs to achieve seamless interaction and collaboration between humans and machines. LLM-HSM comprises four modules: HITL, cyber-physical system (CPS), LLM, and verification, validation, and uncertainty management (VV&UM). Unlike traditional AI-driven automation approaches, LLM-HSM prioritizes human adaptability, real-time decision assistance, and structured validation mechanisms to mitigate risks associated with AI-driven decisions. Prior research on AI applications has largely been fragmented, focusing on specific areas in manufacturing [28]. However, these studies rarely address integration at the system level or the critical dimensions of human involvement and trust. Similarly, earlier approaches to AI-enabled manufacturing quality control often relied heavily on knowledge-based systems that operated from internal records, limiting adaptability and scalability [29]. By contrast, LLM-HSM bridges these gaps by connecting generalized language models with domain-specific manufacturing requirements, enabling context-aware and actionable solutions that extend beyond isolated applications.

Additionally, a key motivation for this study is to encourage the future workforce to explore manufacturing through the use of LLMs, AI, DT technology, and immersive environments. Given the increasing diversity of the workforce, including capability-lessening and disabled workers [21], LLMs can serve as powerful assistive tools to maintain efficiency. Furthermore, LLM-HSM offers significant benefits for novice, inexperienced, and disabled operators, enabling them to interact more effectively with smart manufacturing systems.

Recent studies have begun to explore the application of LLMs in manufacturing [30] and presented conceptual frameworks and taxonomies, especially for developing task-oriented conversational agents in manufacturing [31]. However, these efforts are typically narrow in scope and lack a unified, scalable integration perspective. Most approaches do not adequately address key challenges such as real-time validation, human oversight, interoperability with CPS, and uncertainty management [32–35]. Additionally, the current literature primarily focuses on one aspect of the manufacturing cycle, lacking a holistic approach. For instance, Wang et al. [36] reviewed the evolution and current state of AI-enabled product design (AIPD), highlighting its achievements and challenges, and Fang et al. [37] provided a comprehensive review of context-aware cognitive augmented reality (CA-CAR) assembly. However, a perspective on manufacturing through the lens of GenAI, more specifically LLM, with human-centricity considerations, is lacking. In this study, the LLM-HSM structure is not intended as a validated framework or deployable architecture. Rather, it represents a structured synthesis derived from the reviewed literature and an exemplar integration pattern illustrating how LLMs, HITL, CPS, and VV&UM may interact in smart manufacturing environments. Its purpose is to organize existing research and guide future empirical investigations.

The remainder of the paper is as follows: Section 2 presents a background and a review of the concepts of Industry 5.0, HITL, and LLM; Section 3 reviews the methodology and contributions; Section 4 focuses on current applications of LLM in industry and manufacturing processes, Section 5 presents an exemplar conceptual structure to integrate LLM into HSM; Section 6 discusses the key-enabling technologies and features to realize the concept of LLM-HSM; Section 7 presents the discussions, and Section 8 outlines the future work.

2. Background

2.1. Industry 5.0 and Human-in-the-loop

The evolution of manufacturing systems has been marked by significant paradigm shifts driven by technological advancements and changing production requirements [38]. While Industry 4.0 established the foundation for smart manufacturing through digitalization and automation, Industry 5.0 represents a fundamental shift towards a more human-centric paradigm for manufacturing excellence. Industry 5.0 distinguishes itself by emphasizing the synergistic relationship between human workers and advanced technologies. While Industry 5.0 is widely promoted as a human-centric evolution of industrial systems, much of the existing literature remains conceptual, with limited exploration of how advanced AI, particularly LLMs, can be concretely embedded within collaborative human-machine frameworks to support real-time production intelligence [39].

The key characteristics of Industry 5.0 include human-machine collaboration, which involves integrating human cognitive abilities with technological advancements, sustainable training approaches for workforce development, and enhanced decision-making through the combination of human expertise and artificial intelligence [40]. Technological integration further supports Industry 5.0 through advanced smart technologies that assist human operators, paradigm shifts in planning and developing manufacturing processes, and the establishment of sustainable and resilient manufacturing systems [41]. The emergence of intelligent manufacturing systems, driven by new computational methods such as artificial intelligence and machine learning, has transformed traditional statistical analysis approaches in manufacturing. This transformation facilitates real-time decision support systems, predictive maintenance capabilities, and process optimization through HITL feedback.

Industry 5.0 also emphasizes sustainability and resilience by fostering environmental consciousness in manufacturing processes, promoting social responsibility and worker well-being, and implementing long-term sustainable production strategies. This evolution towards Industry 5.0 represents a crucial stepping stone in manufacturing history, where the focus shifts from pure automation and efficiency to a more balanced framework that values human expertise while leveraging advanced technologies [42]. This paradigm lays the foundation for integrating emerging technologies, such as LLMs, into manufacturing processes, which will be discussed in the subsequent section.

Human-robot collaboration (HRC) emerges as a cornerstone of Industry 5.0, driving the development of robots with enhanced natural interaction capabilities [43]. HRC environments require robots that can seamlessly interact with human operators, understand their intentions, and respond effectively to dynamic situations. This demands advancements in robot perception, cognition, and action intelligence, allowing for intuitive and safe collaboration between humans and robots. The need for intuitive human-machine interfaces (HMIs) further underscores the importance of natural language interaction in HRC [44]. Traditional HMIs often rely on complex programming languages or specialized interfaces, hindering seamless communication between operators and robots. Voice-enabled digital intelligent assistants (DIAs) leveraging natural language processing have the potential to bridge this gap, facilitating a more intuitive and user-friendly interface for human operators in complex assembly tasks [45].

Policy documents from the European Commission explicitly frame Industry 5.0 as a complement to the Industry 4.0 paradigm, in which research and innovation drive a transition towards a sustainable, human-centric and resilient European industry, rather than a purely technology- or productivity-driven evolution [46,47]. These documents emphasize a shift from shareholder value towards stakeholder value, the wellbeing and empowerment of workers, and operating within planetary boundaries as central objectives of future industrial systems. More recently, the ERA industrial technologies roadmap on human-centric research and innovation for the manufacturing sector further operationalizes this vision by detailing how industrial innovation ecosystem stakeholders can achieve human-centric outcomes in technology development and adoption, such as improving workers' safety and wellbeing, upskilling and continuous learning, and user-driven design of artificial intelligence and virtual environments [48]. Taken together, these policy perspectives underpin our interpretation of Industry 5.0 as a socio-technical transition in which digital technologies, including LLMs, are evaluated not only by their contribution to efficiency and automation, but also by how they enhance human agency, safety, and sustainability in smart manufacturing.

In this paper, HITL is adopted as a system-level paradigm, referring to the explicit and continuous involvement of human actors in the perception, decision-making, and control loops of intelligent manufacturing systems [49]. Within the context of Industry 5.0, HITL extends beyond physical interaction to encompass human engagement with cyber-physical systems, decision support tools, digital twins, and AI-driven automation throughout the entire manufacturing lifecycle. HRC is treated as a specific and prominent instantiation of the HITL concept, focusing on physical and cognitive collaboration between humans and robotic systems operating in shared workspaces [50]. While HRC represents a critical enabler of human-centric manufacturing, particularly in tasks that require close human-robot coordination, it constitutes only one subset of the broader HITL framework, which also includes non-robotic human-AI interactions relevant to planning, monitoring, supervision, and validation in smart manufacturing environments [51]. This distinction clarifies the conceptual scope of HITL adopted in this study and ensures consistency with the proposed LLM HSM structure.

2.2. Large language models

2.2.1. Evolution of LLMs

The trajectory of LLMs reflects a significant transformation from early rule-based systems to sophisticated models capable of understanding and generating human-like language with remarkable fluency. The evolution of LLMs is marked by several key phases and elaborated below and shown in Fig. 1. This timeline demonstrates a progression from rule-based and statistical models to sophisticated neural network-based models, with the transformer architecture marking a significant turning point in the development of LLMs.

- *1940s:* Artificial neural networks (ANNs) were introduced, laying the groundwork for future developments in language models [52].
- *1950s-1960s:* The first language models emerged using early neural networks and rule-based systems that relied on linguistic rules and features [53].
- *1980s-1990s:* Statistical language models gained prominence, utilizing probabilistic techniques to capture patterns in language data, building upon earlier models [54]. These models included n-gram models, Hidden Markov Models (HMMs) [55], and Gaussian mixture models (GMMs) [56].
- *Mid-2000s:* Word embeddings were introduced, representing words in a vector space and capturing semantic relationships. This contributed to advancements in natural language processing (NLP).
- *2010:* The recurrent neural network language model (RNNLM) [57], the first neural language model, was introduced to capture sequential dependencies in text and improve contextual understanding [54].
- *Mid-2010s:* Neural language models emerged, utilizing deep learning to comprehend language patterns from vast amounts of text.
- *2015:* Google introduced the Google Neural Machine Translation (GNMT) model [58], marking a milestone in machine translation.

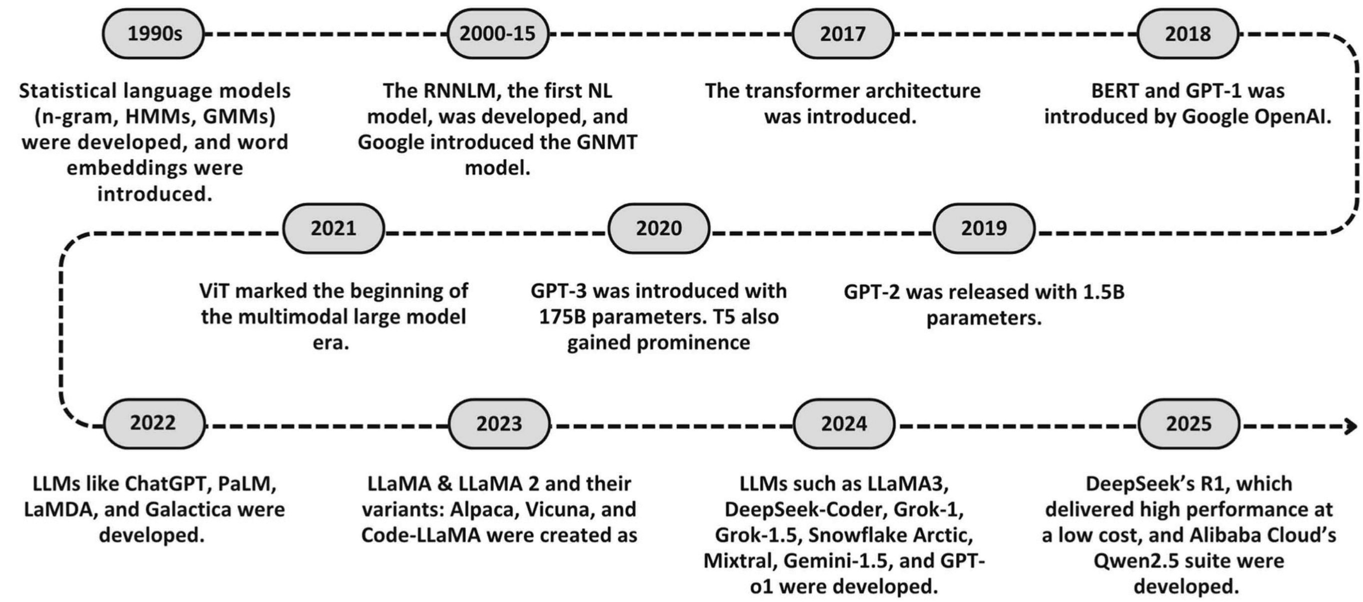

Fig. 1. Evolution of LLMs (ANN: Artificial Neural Network, HMMs: Hidden Markov Models, GMM: Gaussian mixture models, RNNLM: recurrent neural network language model, NL: Natural language, GNMT: Google Neural Machine Translation, BERT: Bidirectional Encoder Representations from Transformers, GPT: Generative Pre-trained Transformer, ViT: Visual Transformer).

- *2017:* The transformer architecture was introduced, enabling parallelization and efficient handling of long-range dependencies. This architecture became the basis for modern LLMs [59].
- *2018: Google introduced* BERT (Bidirectional Encoder Representations from Transformers) [60], enhancing contextual understanding through deep bidirectional representations. OpenAI released GPT-1 (Generative Pre-trained Transformer), the first in the GPT series, showcasing the potential of transformers in NLP.
- *2019:* The Transformer model had a parameter size of 0.165B. GPT-2 was released with 1.5B parameters [61].
- *2020:* GPT-3 was introduced with 175B parameters, demonstrating the emergent capabilities of LLMs. Other models, such as T5, also gained prominence [62].
- *2021:* The proposal of the Visual Transformer (ViT) [63] marked the beginning of the multimodal large model era.
- *2022:* ChatGPT was launched with 175B parameters [64], gaining widespread attention and demonstrating advanced conversational abilities. LLMs like PaLM [65], LaMDA [66], and Galactica [67] were developed.
- *2023:* Meta released LLaMA [68] and LLaMA 2 [69]. Models such as Alpaca [70], Vicuna [71], and Code-LLaMA [72] were created as variants of LLaMA.
- *2024:* LLaMA3 [73] model with 405B parameters. Other LLMs such as DeepSeek-Coder [74], Grok [75], Snowflake Arctic [76], Gemini-1.5 [77], and GPT-o1 [78] were developed.
- *2025:* Significant advancements in LLMs were marked by the release of the DeepSeek R1 model [79], which delivered high performance at a low cost, and Alibaba Cloud's Qwen2.5 suite [80], which supports 29 languages and scales up to 72 billion parameters.

LLMs have evolved from early rule-based and statistical approaches toward deep learning methods, with the transformer architecture marking a decisive breakthrough. For manufacturing systems, this evolution is significant because transformers enable contextual understanding, scalability, and domain adaptation, which were not achievable with earlier models. Recent advances, such as instruction fine-tuning and retrieval-augmented generation, make LLMs capable of handling manufacturing-specific tasks, including interpreting technical manuals, generating machine instructions, and supporting human operators in real-time [81].

2.2.2. LLM and smart manufacturing

LLMs, trained on massive text datasets, represent a significant leap in artificial intelligence, exhibiting impressive capabilities in language understanding, reasoning, and code generation [82]. These advancements open up a vast array of potential applications for LLMs within the realm of smart manufacturing, pushing the boundaries of human-robot collaboration and driving the realization of Industry 5.0 principles. Section 4 will present a detailed elaboration on the current applications of LLM in industry and manufacturing.

LLMs can empower robots with improved natural language comprehension and the ability to interpret complex instructions from human operators, facilitating a more intuitive and seamless HRC experience [44]. Beyond simple command interpretation, LLMs can also be integrated into knowledge-based systems, providing operators with access to vast repositories of technical documentation and best practices. [29]. This enables efficient troubleshooting, knowledge sharing, and enhanced decision-making in complex industrial scenarios.

In addition, LLMs have the potential to address several challenges currently hindering the widespread adoption of Industry 4.0 and 5.0 principles. These many industries often operate with limited financial and technical resources, making it difficult for them to adopt large-scale automation or advanced AI systems [83]. For this reason, affordable, adaptable, and human-centered solutions are essential for enabling them to remain competitive while embracing sustainability and resilience. LLMs can bridge this gap by providing expert-level insights, guiding operators through complex data analysis processes, and assisting in the implementation of AI-driven solutions [82]. Furthermore, LLM can assist in operator training, onboarding new employees, and disseminating critical knowledge throughout the workforce [84]. Although models like GPT-4 and 5, LLaMA, and PaLM have demonstrated impressive language generation capabilities, their industrial adoption remains limited by challenges such as domain adaptation, lack of validation mechanisms, and difficulty integrating with real-time control systems [85,86], issues that current literature rarely addresses in depth. The applications of LLMs in industry will be discussed in further detail in Section 4.

Recent works in LLM-enabled manufacturing show a clear trend toward more ambitious integration of generative AI with digital twins, process planning, and adaptive decision support. In addition to the applications mentioned above, architectural proposals and reviews have also been investigated. For example, Ma et al. [87] introduce a vision and language cobot navigation approach that combines LLMs with CPS infrastructures to enhance human–robot coordination. Similarly, Zhao et al. [88] provide a comprehensive review of foundational AI models tailored for manufacturing, indicating increasing interest in domain-specific large models. Despite these advances, limitations persist: many systems remain task- or domain-specific with minimal mechanisms for uncertainty estimation, verification, or real-time human oversight. There is also little consistency in how inclusivity, verifiable outputs, and human adaptability are handled, especially when models are deployed in dynamic, safety-critical manufacturing environments.

In light of these observations and the previously discussed applications of LLMs in the industry, the literature suggests that simply combining LLMs with CPS or using them as standalone decision agents is insufficient. The gaps exposed by recent studies, such as hallucination in high-precision applications (e.g., aerospace manufacturing) and overconfidence in generative outputs when context is unclear, underscore the need for a formal integration of modules that ensures reliability, human oversight, and validation. The recent efforts, such as LLM-MANUF [89], show movement toward richer frameworks, but even this work only partially addresses uncertainty management and human-centricity. Therefore, to elevate manufacturing systems beyond a descriptive or narrowly focused implementation, the incorporation of VV&UM, along with tightly coupled HITL and CPS modules, becomes essential. The synthesized exemplar LLM-HSM structure aims precisely to fill these integration gaps, offering a more holistic, measurable, and human-aligned model for next-generation smart manufacturing.

2.2.3. LLM and manufacturing-specific key performance indicators

LLMs are not only generic intelligent assistants; they can be explicitly aligned with manufacturing Key Performance Indicators (KPIs) across the product and process lifecycle. In design and engineering, generative AI copilots built on historical product and simulation data have already been reported to cut design time, directly improving KPIs such as design lead time, number of design iterations, and right-first-time design rates. In production planning and scheduling, LLM-enhanced schedulers for human-robot collaborative job shops have demonstrated reductions in average makespan and improved responsiveness to disturbances, resulting in better schedule adherence, lower work-in-progress, and reduced tardiness compared with classical heuristics and DRL baselines [90]. On the shop floor, multimodal LLM and GenAI systems supporting computer numerical control (CNC) programming and process optimization have been associated with shorter setup and programming times, lower cycle times, and fewer execution errors, thereby improving throughput, first-pass yield, and scrap or rework rates [91]. In monitoring and maintenance, LLM-based predictive maintenance frameworks in real industrial plants have achieved high anomaly-detection performance and reported double-digit reductions in operational costs via lower unplanned downtime and more targeted interventions, directly impacting mean time to detection (MTTD), mean time to repair (MTTR), and overall equipment effectiveness (OEE) [92]. Finally, when used as conversational tutors and on-the-job assistants, LLMs can accelerate operator training and support, contributing to reduced time-to-competence and lower error rates during learning [93]. Section 4 will elaborate on this with a focus on different stages of the manufacturing value chain, from design and planning to production, monitoring, maintenance, and training; their stage-specific roles and interactions.

2.3. Humane and technological characteristics

Human abilities and technological capabilities embody two complementary dimensions of intelligence that, when properly integrated, enable a balanced and adaptive manufacturing ecosystem. In the context of smart manufacturing, the distinction between humans and machines is not merely functional but epistemic; humans reason through experience, intuition, and moral reflection, whereas machines operate through data-driven inference, probabilistic reasoning, and deterministic optimization [94]. Understanding this boundary is essential for designing trustworthy, safe, and explainable collaboration frameworks in cyber-physical production systems [14].

Humans excel at making contextual judgments, engaging in ethical reasoning, and solving creative problems in uncertain or ambiguous situations [95]. These strengths are grounded in tacit knowledge, emotional intelligence, and cultural understanding, all of which allow them to interpret intent and manage unforeseen events. However, human performance is constrained by cognitive load, fatigue, and subjective bias, which limit reliability and consistency under repetitive or high-stress conditions. Conversely, technological agents such as LLMs, CPS, and DT demonstrate exceptional capabilities in data processing, scalability, and precision [96]. They can detect complex patterns, simulate dynamic environments, and perform repetitive tasks with high accuracy, but they lack embodied experience, empathy, and genuine semantic understanding, making them unsuitable for autonomous moral or social decision-making.

To foster a symbiotic relationship, human and machine functions must be aligned along explicit decision boundaries, where each agent operates within its optimal domain of expertise. Humans should remain responsible for defining goals, providing ethical oversight, evaluating context, and engaging in creative reasoning, while machines should manage large-scale optimization, prediction, and data synthesis. Adaptive learning loops, enabled by continuous feedback between human judgment and machine inference, can gradually refine these boundaries. Trust is achieved when human operators can understand, predict, and guide the behavior of AI systems, while the system transparently communicates its confidence, rationale, and limitations.

Table 1 summarizes the key domains of humane and technological characteristics, illustrating their core strengths and inherent limitations. This mapping provides the conceptual foundation for allocating decision authority while ensuring that technological intelligence enhances, rather than replaces, human agency in smart manufacturing.

3. Research methodology and contributions

3.1. Contributions

This study makes the following contributions to the literature on smart manufacturing and Industry 5.0.

1. The paper provides a comprehensive and systematic review of recent applications of LLM in smart manufacturing with an explicit focus on HITL involvement.
2. The study synthesizes cross-cutting patterns, trends, and limitations in the existing literature related to human centricity, trustworthiness, and operational maturity. This analysis identifies gaps that clarify where current research falls short of Industry 5.0 principles.
3. Based on the insights derived from the review, the paper presents an exemplar conceptual structure, referred to as LLM-HSM, which organizes the reviewed applications along four recurring dimensions: HITL, CPS, LLMs, and VV&UM. This structure is not proposed as a validated architecture or a deployable solution, but as a reference model to structure existing work, relate disparate application studies, and support future research on human-centric and trustworthy LLM deployment in manufacturing.
4. To address the identified gaps and realize the exemplar conceptual structure, enabling features and technologies are identified and classified.

This study distinguishes itself from several existing reviews, which

Table 1
Humane and technological characteristics, their strengths and limitations.

Characteristic	Agent Type	Core Strengths	Primary Limitations
Cognitive Processing [95,97,98]	Human	- Intuitive reasoning - Contextual judgment - Abstract thinking - Emotional intelligence	- Limited working memory - Fatigue - Cognitive biases - Slower processing speed
	Technology (LLM, AI)	- Rapid data processing - Pattern recognition - Vast memory recall	- No true understanding - Context-dependent hallucinations - Lacks embodied experience - Fixed knowledge cutoff
Decision Making [99,100]	Human	- Ethical consideration - Social awareness - Flexible judgment in ambiguous situations	- Prone to emotional bias - Inconsistent under stress - Subjective judgment variability
	Technology (CPS)	- Data-driven optimization - Real-time simulation - Rapid scenario analysis - Predictive modeling	- Requires accurate models - Cannot handle undefined scenarios - No moral reasoning
Learning & Adaptation [14,101]	Human	- Few-shot learning - Experiential knowledge - Adaptive to novel situations - Generalization	- Requires time and practice - Limited by biological constraints - Knowledge transfer challenges
	Technology (ML, Robotics)	- Large-scale pattern detection - Requires extensive datasets - Continuous improvement through feedback	- Data requirements - Domain-specific - Poor zero-shot performance in novel contexts
Physical Manipulation [96,102]	Human	- Tactile sensitivity - Fine motor skills and dexterity - Intuitive force control - Adaptability to irregular objects	- Strength limitations and fatigue - Risk of injury - Inconsistent performance under stress
	Technology (Robotics)	- Precision - Repeatability, speed, strength, - 24/7 operation	- Limited flexibility - Lack of true tactile sensation - High programming complexity
Sensory Perception [103–105]	Human	- Multimodal integration - Context-awareness - Empathy - Interpretation of subtle cues	- Subjective interpretation - Attentional limitations, - Limited bandwidth
	Technology (CPS, Sensors)	- Continuous monitoring - High-frequency data acquisition - Multi-sensor fusion	- No semantic understanding - Sensor noise and drift - Requires human validation
Communication & collaboration [106,107]	Human	- Pragmatics (sarcasm, irony) - Cultural nuance - Emotional depth	- Slower processing - Ambiguity in expression - Fatigue affects communication
	Technology (LLM)	- Fluent syntax generation - Multilingual processing - Lacks emotional awareness	- No pragmatic understanding - Cultural and emotional blindness - Generates plausible but false information
Creativity & Innovation [108]	Human	- Original ideas from experience - Disruptive innovation - Metaphorical connections	- Time-intensive - Limited output volume - Individual variability, not scalable
	Technology (LLM, AI)	- Novel combinations from existing data - Pattern-based generation - Lacks experiential depth	- Bounded by training data - No genuine originality - Cannot experience or draw from emotions - Produces derivative content
Ethical Judgment [109–111]	Human	- Moral reasoning - Accountability - Empathy-based decisions - Cultural sensitivity	- Time-consuming deliberation - Cultural and personal biases Inconsistent application across cases
	Technology (AI Systems)	- Rule-based compliance - Inherits training data biases	- No moral agency - Cannot be held accountable - Opaque decision-making - Ethical alignment challenges
Contextual Understanding [99]	Human	- Cultural interpretation - Situational awareness - Reading implicit signals	- Limited perspective scope - Cannot process massive data simultaneously - Subjective interpretation
	Technology (CPS)	- Virtual environment simulation - Physics-based modeling - Real-time mirroring - Limited semantic understanding	- Cannot capture human intent/ motivation - Limited predictive accuracy - No emotional modeling
Error Handling [112–114]	Human	- Intuitive problem-solving - Creative workarounds - Adapts to unforeseen failures	- May require significant time - Learning from experience is gradual - Limited ability to generalize
	Technology (CPS, Automation)	- Predictive maintenance - Error detection at scale - Requires explicit fault models - Struggles with novel errors	- Fast diagnosis for known patterns - No creative problem-solving - Black-box reasoning - Over-reliance risks

prioritize broad application surveys or foundational model characteristics. For instance, the review by Mustapha et al. [81] offers a narrative review of applications across mechanics, product design, and manufacturing, providing a conceptual synthesis of emerging use cases. Similarly, Li et al. [115] provide a comprehensive exploration of LLMs' applications in areas such as quality control, supply chain management, and robotics, focusing on insights and challenges. While these papers successfully establish the wide-ranging potential of LLMs in manufacturing, the present research takes a crucial next step by synthesizing a modular architecture to structure existing work. By making VV&UM and HITL core, dedicated modules, we address the trustworthiness and contextual accuracy gaps often highlighted in broader application surveys [81,115].

LLM-HSM also compares and extends reviews focused on specific integrations or manufacturing philosophies. Reviews such as Wu et al. [116] focus on the integration of Multimodal LLMs (MLLMs) and embodied intelligence to enhance HRC within smart manufacturing. Chen et al. [32] specifically investigate the LLM and DT integration in Industry 5.0, proposing the Interactive-DT framework, which addresses the challenges of interoperability and data management. The current study builds on these foundational concepts by encompassing CPS (which includes DTs), but crucially embeds this technological stack within the overriding requirements of human-centric deployment (HITL) and reliability assurance (VV&UM). Meanwhile, Oyekan et al. [117] offer decision-making guidance for selecting between traditional NLP methods, Knowledge Graphs, and LLMs based on experimental evaluation and domain dependency. In contrast, we assume the adoption of LLMs and focus on solving the deployment challenge, rather than model selection, by prescribing a framework that ensures compliance and performance for high-precision environments.

Furthermore, the current study differentiates itself from reviews with highly overlapping thematic scopes, such as those focusing on Industrial Foundation Models (IFMs) or New-Generation Intelligent Manufacturing (New-IM). Zhao et al. [88] conduct a systematic review on IFMs, summarizing applications and exploring key technologies like Retrieval-Augmented Generation (RAG) and Knowledge Graphs (KG). Zhang et al. [118] similarly provide a survey on the pathways and challenges of LLMs in New-IM toward Industry 5.0, proposing the New-IM-LLM theoretical framework. While these works align strongly with Industry 5.0 and intelligent manufacturing, they generally lack the verification of AI outputs. In this study, we address the limitations (such as performance uncertainty and trustworthiness) identified in related studies by dedicating modules to VV&UM and HITL, providing a structured approach to integrate barriers, manage large amounts of data, and ensure compliant outputs.

In addition, the review by Ma et al. [87] comprehensively maps LLM applications across six manufacturing domains and synthesizes the synergies between LLMs and various Industry 5.0 enabling technologies, such as DTs and industrial internet of things (IIoT). Similarly, the survey by Zhang et al. [119] focuses on establishing the core advantages of Large Scale Foundation Models (LSFMs) in terms of generalization and automated dataset generation, systematically comparing these capabilities against the recognized shortcomings of traditional deep learning methods in areas like Prognostics and Health Management (PHM) and quality control. In contrast to these papers, which primarily review current status and potential, the current study builds upon these established application areas to synthesize a concrete architecture that explicitly mandates human oversight and reliability checks to mitigate risks like hallucination and geometric reasoning deficiencies that the other reviews identify as key challenges

Together, these studies highlight the multidimensional capabilities of LLMs and the evolving synergies between LLMs and smart manufacturing applications, paving the way for future research on LLM-HSM. Table 2 provides a comparison of existing surveys in LLM and manufacturing, integrating insights from these recent contributions. It is worth mentioning that the review papers in the following table were extracted outside the PRISMA approach, which will be introduced in Section 3.2.

Table 2
Comparison of existing review papers.

Research	Year	Approach	Application areas	Framework/ Architecture/ Structure	Implementation	Contributions
Zhang et al. [119]	2023	Systematic survey with comparative analysis, technical roadmaps	PHM, Quality control, HAR, Industrial robotics	—	Case studies on PCB inspection & human action recognition	Roadmaps for LSFMs in manufacturing
Li et al. [115]	2024	Review of applications, evaluation of the performance of LLMs	Quality control, Supply chain, Robotics, Knowledge management	—	Case studies using GPT-4V for text processing, code generation, and VR development	LLM application insights, challenges, and future directions for manufacturing advancements
Zhang et al. [118]	2025	Cross-disciplinary analysis and insights	Design, Production, Service	New-IM-LLM	—	Cross-disciplinary analysis, new-generation IM-LLM architecture
Mustapha [81]	2025	Narrative review of recent literature and use cases	Mechanics, Product design, Manufacturing	—	Conceptual synthesis with illustrative examples	Identifies opportunities and limitations of LLMs across engineering workflows
Oyekan et. al. [117]	2025	Comparative analysis with experimental evaluation	Manufacturing robotics, HRC, Industry 5.0 contexts	LKLM	Includes experimental evaluation and metrics	Provides decision guidance, sector-specific insights, and ethics considerations
Chen et al. [32]	2025	Thematic review with a proposed three-layer framework	Digital twin construction and operation, intelligent manufacturing	Interactive-DT	—	Clarifies LLM roles in DT pipelines and highlights open challenges
Ma et al. [87]	2025	Review, synthesis of synergies & enabling technologies	Design, planning, execution, quality control, supply chain	—	—	Comprehensive mapping of LLM applications and enablers in Industry 5.0
Zhao et al. [88]	2025	Literature review on IFMs and enabling techniques	Manufacturing knowledge bases, predictive maintenance, process optimization, and human-machine collaboration	IFM	Examples from predictive maintenance, process knowledge extraction, and digital assistants	Provides an overview of IFMs in manufacturing, focusing on RAG, embeddings, & KG
Wu et al. [116]	2026	Structured review around the Perception-Cognition-Actuation loop	Human-robot collaboration, smart manufacturing	MLLM	Conceptual only, with discussion of datasets and platforms	First to integrate HRC, MLLMs, and spatial intelligence into one review
This study		Comprehensive review focusing on applications and implementations with a framework	Design & Simulation, Planning & Scheduling, Production, Monitoring & Control, Repair & Maintenance, Training & Education	LLM-HSM	Exemplar conceptual structure	Comprehensive review of applications, HITL perspective, exemplar LLM-HSM structure, identifying technologies, and gap analysis.

HAR: Human Action Recognition, PCB: Printed Circuit Board, VR: Virtual Reality, LKLM: Large Knowledge Language Model,

3.2. Methodology

To realize the contributions outlined in Section 3.1, first, we reviewed the current applications of LLM in smart manufacturing through a systematic scoping review following the updated Preferred Reporting Items for Systematic Reviews (PRISMA 2020) guideline [120]. The adopted protocol comprised the definition of information sources, formulation of search strategies, specification of eligibility criteria, and structured processes for selection and data collection. The PRISMA flow diagram (Fig. 2) illustrates the overall procedure. More details of the PRISMA approach are outlined in Sections 3.1.1–3.1.3.

Based on the findings and gap analysis performed in Section 4, we integrated LLMs into the already-developed concept of H-SM [21] to synthesize the LLM-HSM exemplar conceptual structure. Then, we identified the enabling features and technologies that support its realization. Finally, the study discusses the relevance to Industry 5.0, socio-technical and organizational implications of LLM-HSM, ethical and human rights safeguards in LLM-HSM, and the challenges ahead.

3.2.1. Information sources and search strategy

The literature search was conducted across five well-recognized databases: Scopus (n = 181), ScienceDirect (n = 382), Web of Science (n = 79), IEEE Xplore Digital Library (n = 193), and SpringerLink (n = 248), resulting in 1086 initial records. The search strings combined keywords relating to smart manufacturing, Industry 4.0/5.0, and human-in-the-loop concepts with terminology connected to LLMs and generative AI. The applied string was: ("smart manufacturing" OR "industry 5" OR "industry 4") AND ("LLM" OR "large language model" OR "genai" OR "generative AI" OR "generative artificial intelligence") AND (human OR "operator"). The search was performed in the title, abstract, and keyword fields. Filters were applied to restrict results to English-language publications from 2017 to 2026, while excluding conference papers, reviews, and editorials to ensure consistency and quality.

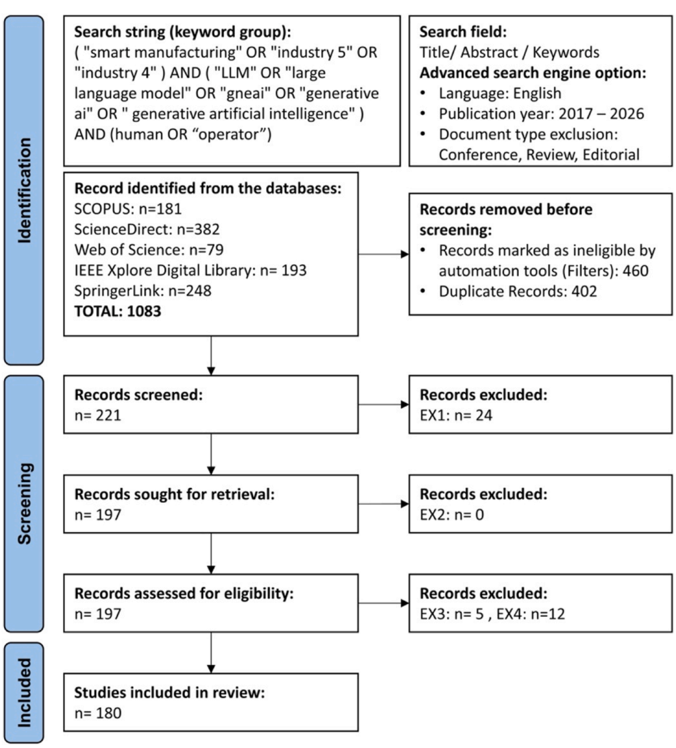

Fig. 2. PRISMA flow diagram illustrating the identification, screening, eligibility, and inclusion process for studies in the systematic scoping review.

3.2.2. Eligibility criteria

The inclusion criteria were defined to ensure the review captured the most relevant and rigorous contributions. Studies were eligible if they addressed the application or potential of LLMs in manufacturing, Industry 4.0/5.0, or human-in-the-loop systems, if they reported empirical, conceptual, or technical insights with relevance to smart manufacturing, and if they were peer-reviewed journal articles or high-quality indexed publications in English. Exclusion criteria were applied at different stages: (EX1) studies outside the defined thematic scope, (EX2) inaccessible or unretrievable full texts, (EX3) publications with insufficient methodological or contextual detail, and (EX4) duplicates that were not automatically filtered.

3.2.3. Selection and data collection process

From the 1086 records initially retrieved, 460 were removed by automation tools based on filters, and 402 were identified as duplicates, resulting in 224 records for screening. Title and abstract screening excluded 24 records (EX1) due to irrelevant content or duplication that were not identified in the previous process, resulting in 197 studies that were sought for retrieval. No studies were excluded due to inaccessibility (EX2 = 0). During the eligibility assessment, five records were excluded for limited methodological or contextual relevance (EX3), and twelve records were excluded for insufficient detail or scope (EX4). Ultimately, 180 studies were included in the final review dataset.

The selection process was conducted by multiple reviewers to minimize bias, with disagreements resolved through a consensus-based approach. Relevant data were extracted using a standardized form capturing bibliographic information, research focus, methodology, and contributions. The final dataset served as the foundation for analyzing trends, enabling the exploration of technologies, case studies, and scenarios related to LLM-HSM.

The screening process involves a specific set of keywords. Fig. 3 shows the overlay visualization of the keyword occurrences, generated by VOSviewer, in LLMs and smart manufacturing areas. The most used keywords include large language model, smart manufacturing, natural language processing, computational linguistics, artificial intelligence, machine learning, and chatbot.

4. Current applications of LLM in smart manufacturing

This section reviews the current applications of LLMs in manufacturing. For this, we categorized these applications with respect to different industrial sectors, including smart factory, robots, conventional manufacturing processes, and AM processes. The proposed paradigm can be applied in various fields, including design, simulation and scheduling, production, monitoring and control, repair and maintenance, and training and education. A considerable number of papers have made efforts to integrate LLMs with the manufacturing processes. For instance, Rane et al. [121] explored the contributions of ChatGPT in manufacturing across various fields, including process optimization, quality control, predictive maintenance, supply chain management, human-machine collaboration, knowledge transfer and training, fault diagnosis, decision support, customization and personalization, and cybersecurity.

To carry out this classification, among the 180 papers selected using the approach in Section 2, we performed a full-text screening of the papers to extract those that applied LLM or GenAI in a manufacturing setting. Of these, 76 papers met this inclusion criterion. Table 3 demonstrates the results of this classification. The last column in the table also summarizes the observed trend for each application, which will be thoroughly discussed in Section 4.7. Each application field will be elaborated on in the following section, and the respective published papers will be discussed.

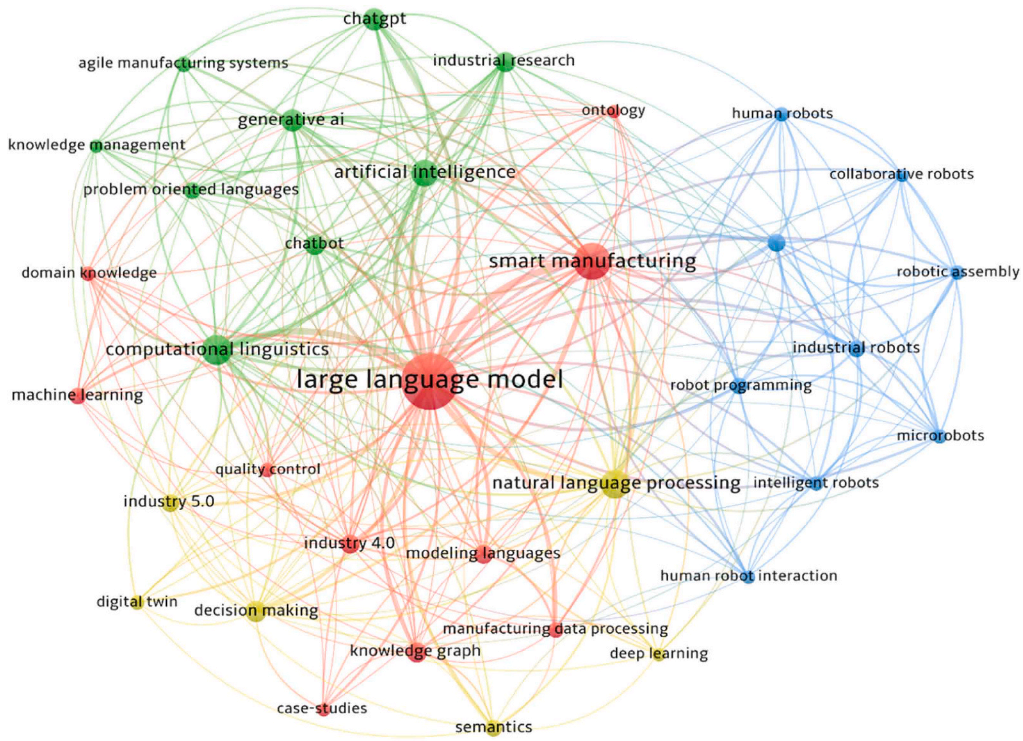

Fig. 3. Overlay visualization of the keyword occurrence on “large language model” and “smart manufacturing” from the Scopus database generated by VOSviewer.

Table 3
Current applications of LLMs in various industrial sectors and the observed trends.

Applications	Industrial Sectors				Counts	Observed Trend
	Smart Factory	Robot	AM Process	UM Process		
Design and Simulation	[122]	–	[123,124]	[125–132]	11	- Expanding use in DfM and CAD assistance. - Text-to-image/3D pipelines emerging in AM. - Most remain prototypes; risk of hallucination.
Planning and Scheduling	[133–136]	[137–144]	[145]	[146,147]	15	- Strong growth in LLM-enabled multi-agent planning. - Use of RAG to build taxonomies. - Real-time scaling and uncertainty handling still limited.
Production	[148–150]	[44,151–154]	[155]	[45,156–159]	14	- Active in G-code synthesis and HRC. - LLM agents orchestrate modular tasks. - Needs CPS/DT grounding for safe execution.
Monitoring and Control	[160–164]	[165–171]	[172–174]	[29,175–180]	21	- Most active category overall. - Growth in KG + LLM fault diagnosis. - Sensor/DT fusion emerging. - Online verification still inconsistent.
Repair and Maintenance	[181–185]	–	–	[186–188]	8	- Underexplored despite high potential. - Ontology-guided fine-tuning used. - Few validated field studies.
Training and Education	[189,190]	[191]	[192,193]	[194]	6	- Early-stage adoption. - AR/VLM + DT training shows promise. - Inclusivity and outcome evaluation limited.

4.1. Design and simulation

LLMs can play a significant role in design for manufacturing (DfM) and simulation processes, improving the creation of parts and assemblies. They act as a repository of manufacturing expertise, utilizing their abilities in pattern recognition and language comprehension throughout both design and production. Their capabilities in programming and text analysis allow them to generate and modify design and manufacturing documents. Historically, DfM has relied on human knowledge and CAD software to evaluate and enhance designs for improved manufacturability. Incorporating LLMs into this workflow can enhance the efficiency of DfM, resulting in more consistent, scalable, and effective decision-making.

There are numerous applications for domain-specific LLMs in DfM, as shown in Fig. 4. In this study [122], first, LLMs can assist in determining the most suitable manufacturing methods for a part based on its characteristics. They can also suggest and implement design modifications to enhance manufacturability and optimize production workflows. This can extend to part sourcing, where the model’s reasoning capabilities help select appropriate suppliers based on the part’s function and performance criteria. Lastly, LLMs could support the generation of manufacturing instructions for various processes [122].

AI, capable of generating images from text input, has recently gained significant public attention [132]. Like AI chatbots, Text-to-image

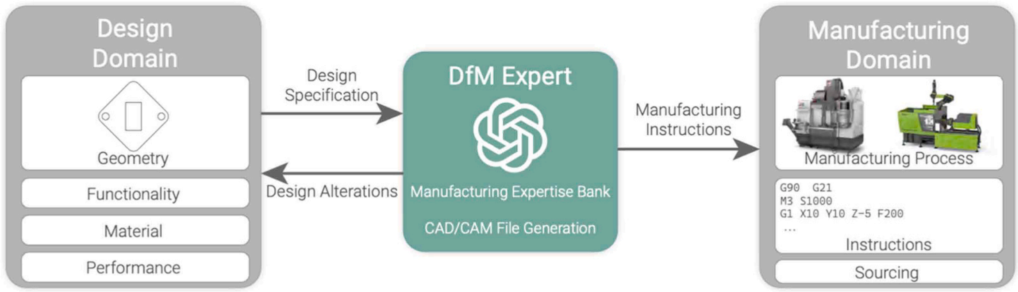

Fig. 4. Integration of GPT-4 into Design for Manufacturing (DfM). GPT-4 can be used to select optimal manufacturing techniques based on a part's features. Furthermore, it can propose and implement modifications to a design to improve its manufacturability, ultimately leading to more efficient production processes. Adapted from [122].

algorithms are mostly utilized in AM research, though their use remains limited [124]. Currently, no industrial applications or solutions are available. Tools like Midjourney are increasingly being employed to streamline design tasks before the printing phase or to generate initial design concepts for print files from text prompts [128]. These text inputs can also come from chatbots like ChatGPT, with the resulting 2D images serving as inputs for generating realistic 3D models [129]. The emergence of Text-to-image algorithms has also sparked greater interest in generative algorithms for creating 3D objects [130]. Recent progress in pre-trained Text-to-image diffusion models has led to the creation of text-to-3D algorithms capable of producing realistic 3D or CAD models from text prompts or 2D images [131].

4.2. Planning and scheduling

LLMs improve planning and scheduling by analyzing complex datasets to develop optimized strategies and actionable insights. They can dynamically prioritize tasks, allocate resources, and simulate various scenarios, facilitating real-time adaptive decision-making for enhanced operational efficiency. Several studies have explored the applications of LLMs in smart factories [133–136], robotics [137–144], AM processes [145], and UM processes [146,147]. For instance, the work by Lim et al. [136] demonstrates the use of an LLM within a multi-agent framework to improve the responsiveness of manufacturing processes. As shown in Fig. 5, their proposed framework starts with a user prompt in natural language describing the desired product goal. An LLM agent then translates this into instructions, which guide other agents, initiating a chain of actions until the task is completed.

Moreover, LLMs are being integrated into DT systems for production planning and control. A novel approach [134] organizes production operations within a hierarchical structure based on the automation pyramid, utilizing LLM agents to interpret production data and generate process plans. This integration creates more intuitive production facilities with enhanced task automation and flexibility. LLMs are also being applied in multi-agent systems to manage production planning and scheduling. As discussed in [135], an LLM-based multi-agent manufacturing system was proposed to tackle challenges in multi-variety and small-batch production. The system utilizes LLMs to evaluate shop-floor conditions and select the most suitable machines for specific tasks. LLM-supported negotiations between agents ensure efficient resource allocation, optimize scheduling, and improve responsiveness to fluctuating production needs.

In another study, the LLM A* framework [139] leverages the commonsense capabilities of LLMs to facilitate few-shot near-optimal path planning. This approach enables more efficient exploration of the search space and yields paths comparable to those of traditional algorithms, while outperforming reinforcement learning methods. In addition, Yoneda et al. [140] introduced Statler, a framework that provides LLMs with a clear representation of the world state through a continuously updated 'memory.' By facilitating access to this world state

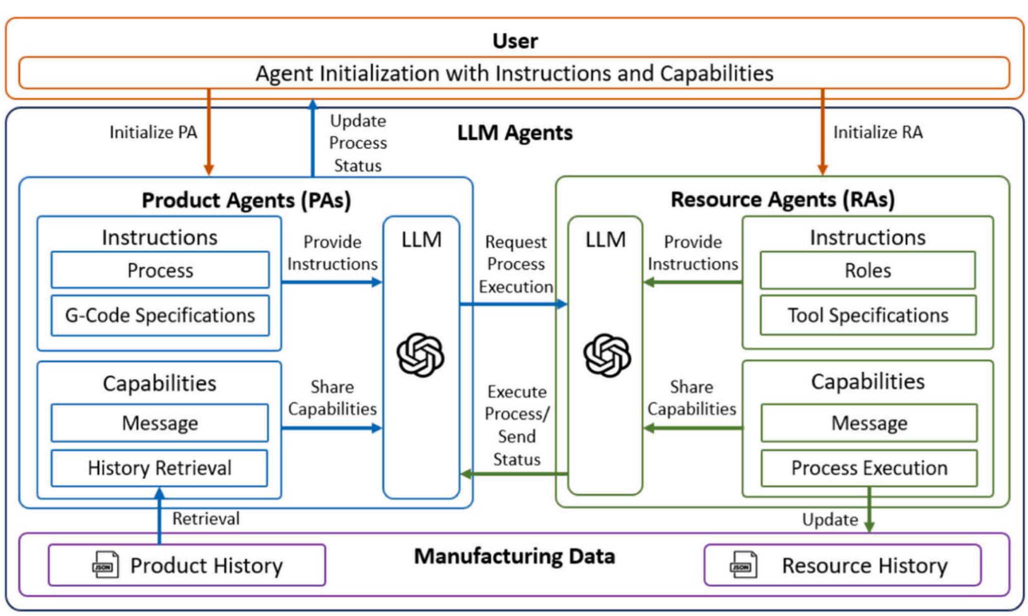

Fig. 5. Framework for LLM-Enabled Multi-Agent Manufacturing Systems. It demonstrates an overview of an LLM-enabled multi-agent manufacturing system framework, where users initialize agents with instructions and capabilities. Adapted from [136].

'memory,' Statler enhances the ability of LLMs to reason about planning tasks that require long time horizons, thereby overcoming limitations related to context length. Similar models, such as EmbodiedGPT [141], grounded decoding (GD) [142], and ProgPrompt [144], have also been developed to help establish plans and execute tasks for robots.

To utilize LLMs in the planning/scheduling of manufacturing processes, Liu et al. [145] proposed an innovative named entity recognition (NER) framework. This framework leverages the advanced capabilities of LLMs to overcome the limitations of manually defined taxonomies. The authors integrated expert knowledge from both academic literature and LLMs through a method called retrieval-augmented generation (RAG). This approach allows for the automatic customization of taxonomies tailored to specific manufacturing processes. In addition, Xu et al. [146] investigated the potential of integrating generative AI, specifically LLMs, into intelligent process planning. The researchers proposed a conceptual framework that combines LLMs and DTs to enhance the planning and scheduling of production processes. LLMs are utilized to model complex scenarios and generate process plans that can be verified and refined in real-time.

4.3. Production

LLMs have become indispensable tools for enhancing automation, streamlining processes, and facilitating informed decision-making in production. By analyzing extensive streams of real-time data from sensors and IoT devices, LLMs provide actionable recommendations for fine-tuning machine parameters, enhancing efficiency, and ensuring consistent product quality. Applications of LLM form production in different industrial sectors, including smart factories [148–150], robotics [44,151–154], AM processes [155], and UM processes [45, 156–159] have been studied.

In the context of smart factories, LLMs are being integrated to enhance various aspects of manufacturing processes. The concept of a "Unified Industrial Large Knowledge Model Framework", as shown in Fig. 6, [149] is proposed to bridge the gap between general LLMs and domain-specific industrial knowledge, potentially revolutionizing Industry 4.0 and smart manufacturing. This framework emphasizes the importance of specialized domain knowledge in effectively addressing complex industrial applications. Another significant application is the development of LLM-based agents that work within modular production systems. For instance, in [150], the authors introduced a framework that combines LLMs with DTs and industrial automation systems. This integration enables intelligent planning and control of production processes, where LLM agents interpret descriptive information in DTs and autonomously orchestrate sequences of tasks.

Moreover, LLMs have been integrated with robotics to improve the HRC in assembly lines [152], assist operators with tool retrieval tasks [44], reduce cognitive workload, and improve efficiency [45]. For instance, in [152], the LLM-based system enhances human-robot collaboration by providing a natural language interface for operators and integrating real-time data through DTs. This allows for more intuitive interactions and reduces the complexity of programming robots, making the system more user-friendly and efficient. The study demonstrated significant improvements in collaboration and efficiency, showcasing the potential of LLMs to streamline operations in smart factories. In [44], a vision and language cobot navigation system was proposed, where LLMs are used in human-centric smart manufacturing to assist operators with tool retrieval tasks. The integration of LLMs allows collaborative robots to understand natural language instructions and perform navigation tasks autonomously, significantly reducing interruptions in the manufacturing process.

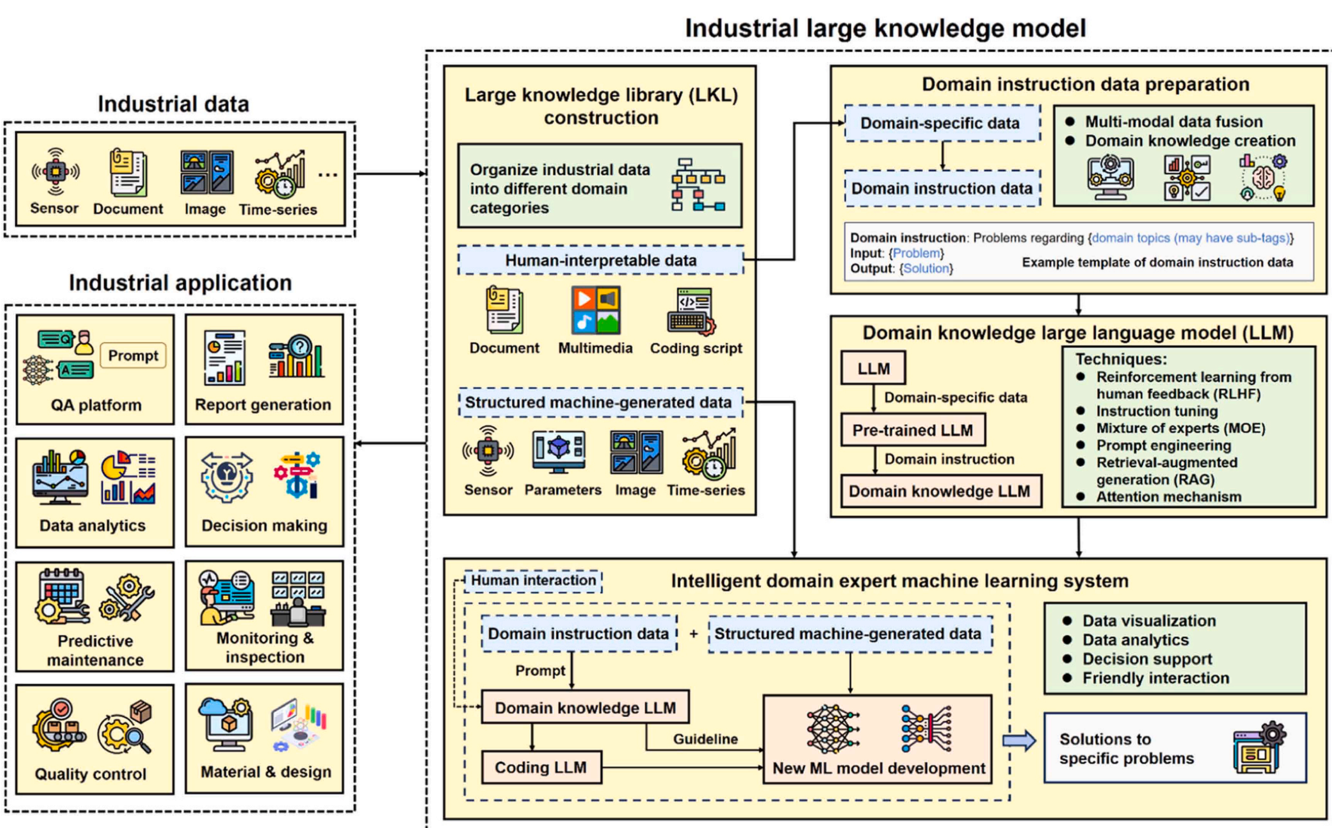

Fig. 6. Proposed unified industrial large knowledge model framework. The domain instruction data are organized into three parts: first, the domain instruction, which identifies the problem's domain and may include sub-tags for refined categorization; second, the in- put, which clearly outlines the current problem; and third, the output, which presents the corresponding solution. Adapted from [149].

4.4. Monitoring and control

LLMs are also being integrated into quality control and fault diagnosis systems to enhance the detection and analysis of defects in manufacturing. Application of LLM in monitoring and control for smart factories [160–164], robotics [165–171], AM processes [172–174], and UM processes [29,175–179]. One notable application is the CausalKGPT model [162], which combines causal knowledge graphs with LLMs to analyze quality-related defects in aerospace product manufacturing. This model, as shown in Fig. 7, improves the ability to identify the root causes of defects and provides a more accurate analysis by leveraging causal associations in quality-related data. In [163], a joint knowledge graph and LLM model were proposed for fault diagnosis in aviation assembly. The model uses graph-structured data within knowledge graphs to improve LLM-based reasoning, enhancing the ability to localize faults and generate troubleshooting solutions. By integrating knowledge graphs with LLMs, the model provides more accurate and efficient fault diagnosis in complex assembly processes. Another noteworthy application is in the area of CPS, as discussed in [164]. LLMs are used to enhance CPS intelligence by providing advanced reasoning and language comprehension capabilities. This integration allows CPS to operate with a higher degree of autonomy, making decisions based on real-time data and improving the overall responsiveness and efficiency of automated manufacturing systems.

LLM has been beneficial to robotics in a number of ways. The ManiGPT paradigm [166] uses LLMs to manipulate robots in an interactive manner. This technology makes human-robot interaction more natural and adaptive by enhancing the responsiveness of robots in helping people with manipulation activities by comprehending human intent through discourse and creating appropriate manipulation plans. Furthermore, Sun et al. [167] looked at the use of LLMs for the all-encompassing control of humanoid robot movement. They demonstrated that LLMs can control humanoid robots with human-level performance by validating their work using university simulations. Research is also being done on employing different implementations of LLM models to provide commands that directly control a robot's actuators. Among these, the Google research team created the groundbreaking RT-1 [168], RT-2 [169], and AutoRT [122] projects, which result in improved performance on both known and unknown tasks. A layered architecture and modular design were used in [171] to translate natural language commands into executable tasks for UAVs (Unmanned Aerial Vehicles) and UGVs (Unmanned Ground Vehicles).

LLMs are also being used to enhance troubleshooting in AM. As explored by Badini et al. [172] and Fan et al. [173], LLM was employed to address common issues in AM, such as bed detachment and warping and the G-code generation process, respectively. Additionally, LLMs are being fine-tuned for quality control tasks through error-assisted learning. An error-assisted fine-tuning technique was presented in [178] to modify LLMs especially for the manufacturing industry. This technique enhances the LLM's capacity to recognize important components inside the domain and produce trustworthy code in answer to manufacturing inquiries. In addition, a recent study by Fan et. al., [180] has demonstrated the application of a vision language model (VLM) in in-situ monitoring and control. They concluded that In anomaly detection and process monitoring, their model, MaViLa, demonstrates improved performance when abstract manufacturing defects are described using concrete, familiar analogies.

4.5. Repair and maintenance

Repair and maintenance processes have been significantly transformed with the adoption of LLMs, which have introduced new paradigms for reducing downtime and optimizing equipment performance. In comparison to other applications, the integration of LLMs with repair and maintenance in the industry has not been given adequate attention and is mostly focused on smart factories [181–185] and UM processes [186–188]. For instance, a domain knowledge base has been created from manuals about troubleshooting, operations, and maintenance expertise, aiding in fault identification and guidance for wind turbines [182]. In a separate study, Wang et al. [195] introduced an intelligent

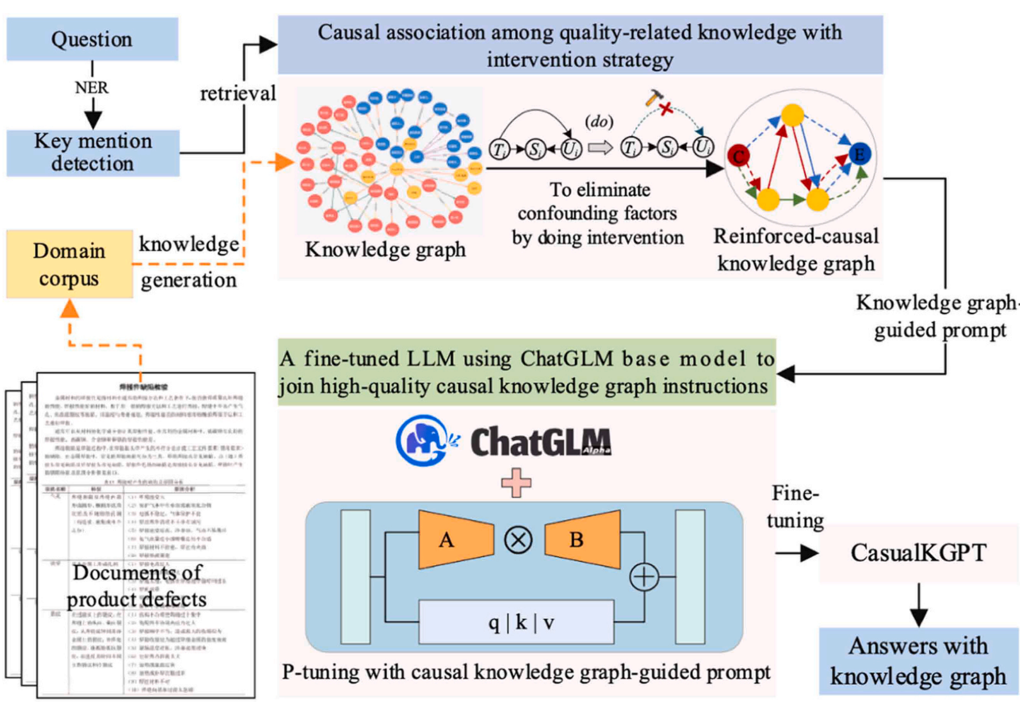

Fig. 7. CasualKGPT aims to integrate causal knowledge graphs with LLMs to analyze quality-related defects in aerospace product manufacturing. The proposed method is mainly divided into two stages. (i) The quality-related corpus data selected from documents in aerospace manufacturing are cleaned. Then, the structure is designed to build a high-quality causal knowledge graph. (ii) On this basis, they fine-tune a LLM by taking ChatGLM as a base model and designing a causal knowledge graph-guided P-tuning strategy. Adapted from [162].

maintenance assistant designed for aircraft service technicians. The research involved a fine-tuned GPT-3.5 model, with the fine-tuning process incorporating an integrated dataset that merges aircraft structure ontology with a curated maintenance log (Fig. 8). The authors noted that the ontology data provided the LLM with domain-specific, hierarchically structured knowledge, which helped to clarify ambiguities related to similar component names across different sub-components of the aircraft.

Additionally, Wang et al. [188] introduced a maintenance framework designed to support intelligent decision-making focused on safety and reducing defective products. Built on an LLM enhanced with proprietary knowledge and a LoRA-finetuned large vision model, the framework was tested in a wafer manufacturing environment, demonstrating outstanding performance in visual problem analysis and solution recommendations. Similarly, [184,185] proposed a semantic search framework for maintenance decision support. The Technical Language Processing (TLP) system uses a fine-tuned BERT model to process complex industrial texts. The authors validated the TLP pipeline's effectiveness with aviation and mining industry-specific maintenance data, achieving search return precision rates of 0.97 and 0.94, respectively.

Without a doubt, LLMs may offer crucial assistance when performing repairs. Thanks to their multimodal capabilities, they can provide detailed repair instructions in various formats, including interactive manuals and annotated photographs. Repairs can be completed effectively since technicians can communicate with the models to ask questions or consider different strategies. Additionally, LLMs serve as archives of organizational knowledge, providing new technicians with access to best practices and previous solutions.

4.6. Training and education

Training and education in industrial settings have undergone a significant transformation with the integration of LLMs. These models offer interactive and personalized learning experiences tailored to the diverse needs of employees across various skill levels. Workers can use LLM-powered systems to acquire new technical skills or deepen their understanding of complex industrial processes. By leveraging advanced natural language interfaces, these systems deliver immediate feedback and clarification, making learning more engaging, efficient, and adaptable. The Integration of LLM with training and education in different industrial sectors, such as smart factories [189,190], robotics [191], AM processes [192,193], and UM processes [194], has been explored.

For instance, LLMs are being used in AM training programs, especially for non-experts. A unique training system that integrates a VLM and AR within a DTs framework was presented in [192]. With the help of visual overlays that walk learners through intricate industrial procedures, this system offers an immersive learning environment. Using LLMs to improve knowledge sharing in manufacturing settings is another example. LLMs were included in a system designed to retrieve data from expert knowledge and manufacturing documentation, as described in [190]. On the manufacturing floor, this technology enables operators to make informed decisions and resolve issues by providing them with rapid access to relevant information. The study also emphasized how crucial it is to keep learning and adapting because user interactions could cause the LLM system to change over time, increasing its effectiveness even more. Additionally, Westphal et al. [196] developed a chatbot that uses Fused Deposition Modelling (FDM) and a dedicated 3D printer to help novice users navigate the 3D printing process. The chatbot begins by offering fundamental knowledge about the process, explains the overall workflow, and then provides step-by-step guidance throughout the printing process, as shown in Fig. 9.

LLM finds great application in many other fields. LLMs, for example, can replicate realistic, scenario-based training courses, therefore providing staff members with a controlled environment in which to practice decision-making and issue resolution. This enhances self-esteem and ensures that children are prepared for real-world situations. Furthermore, LLMs simplify the process of developing tailored training courses designed to meet specific job roles and corporate objectives, facilitating long-term competency development. LLMs create content that fills in knowledge gaps and provide a personalized learning route by examining job descriptions, past performance information, and individual learning tendencies. LLMs also enable the creation of training modules tailored to each function and educational requirement.

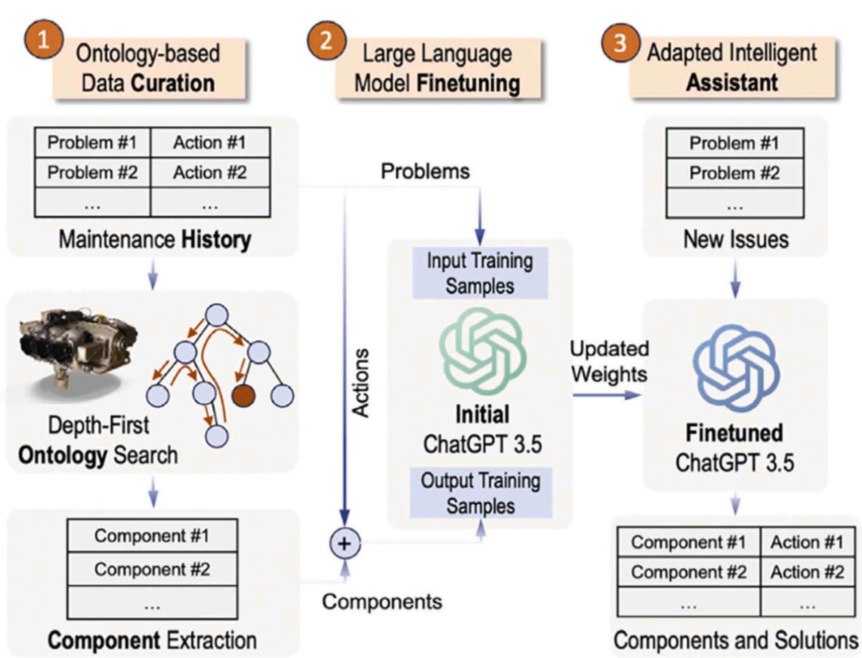

Fig. 8. Ontology-integrated tuning of LLM for predictive maintenance. The ontology of aircraft structure is first investigated to curate the original maintenance logs into conversational data. The conversational data is used as input to fine-tune the GPT-3.5 model, enabling it to achieve a contextual understanding of the components' hierarchical structure. Adapted from [195].

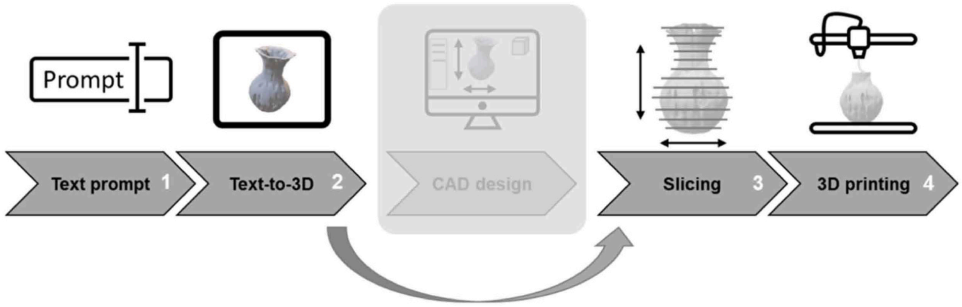

Fig. 9. Process flow from text prompt, via a Text-to-3D algorithm, to the printed component. Adapted from [193].

4.7. Emerging trends and research gaps

Across the 76 papers for the analysis in this Section, activity clusters center on monitoring and control, as well as planning and scheduling, where LLMs are increasingly paired with knowledge graphs, retrieval-augmented generation, and multi-agent patterns to reason over plant states and coordinate tasks. Production exhibits steady growth in code synthesis and HRC but still relies on CPS and DT grounding to ensure executable safety and feasibility. Design and simulation benefit from LLM-assisted DfM and emerging text-to-3D workflows in AM, though most remain prototype-level and sensitive to hallucination without verification [197]. In contrast, repair and maintenance, as well as training and education, are comparatively underexplored, despite their clear potential for reducing downtime and providing operator support. These distributional imbalances, together with the limited and uneven treatment of uncertainty, verification, and human oversight, reinforce the fact that simply combining LLMs with CPS or DTs is insufficient; tangible progress requires their integration within LLM-HSM so that VV&UM and HITL systematically close reliability and inclusivity gaps while scaling to real shop-floor conditions. Several key research gaps persist, limiting the development of intelligent, verifiable, and human-aligned manufacturing systems. These gaps form the foundation for the proposed LLM-HSM framework.

- *Fragmented and task-specific implementations:* Most studies employ LLMs for isolated functions such as design generation, process planning, code synthesis, or fault diagnosis within specific industrial sectors. There is no unified architecture that integrates these functionalities across the full manufacturing lifecycle, from design to maintenance and operator training. Consequently, current approaches fail to capture the interdependencies among perception, cognition, control, and feedback [198]. The absence of an end-to-end structure prevents scalable knowledge transfer between domains and restricts the potential for continuous learning and adaptation.
- *Limited integration with CPS and DT:* Although several studies demonstrate LLM-CPS or LLM-DT integrations [199–201], these are typically ad hoc and lack formalized data flows, interface standards, and control boundaries. As highlighted in Section 4.7, production systems still depend on CPS or DT grounding to ensure executable safety, yet current approaches rarely define how this grounding occurs or how data fidelity is maintained. Without explicit architectural coupling, LLM outputs remain context-sensitive and prone to hallucination or unsafe execution when transferred to physical environments.
- *Deficiencies in VV&UM:* Hallucination, overconfidence, and inconsistent online verification recur across reviewed applications, particularly in high-precision manufacturing. While certain studies incorporate error-assisted fine-tuning [89] or causal reasoning, these mechanisms remain problem-specific and lack generalizability [34]. There is currently no systematic approach for embedding VV&UM processes within LLM-enabled manufacturing architectures. This shortfall limits trust, reproducibility, and regulatory acceptance.
- *Weak modelling of human roles and collaboration mechanisms:* Despite frequent references to human-in-the-loop and human-robot collaboration, the literature seldom formalizes how human judgment interacts with AI reasoning or how feedback is integrated into system learning. Operators typically act as prompt-givers or passive supervisors rather than active collaborators in the decision cycle. Moreover, inclusivity, explainability, and operator adaptability are inconsistently addressed, particularly in training and education contexts. A structured definition of human roles, feedback channels, and adaptive interfaces is required to achieve genuine human-centric manufacturing aligned with Industry 5.0 principles.
- *Lack of benchmarking frameworks and evaluation criteria:* Current research provides valuable prototypes and conceptual models but lacks standardized criteria for assessing cognitive, collaborative, and operational performance. Comparative evaluation of alternative architectures is rarely feasible due to the heterogeneity of methods and datasets. As noted by several recent relevant works [32,199,202,203], research efforts increasingly highlight this methodological gap, emphasizing the need for reproducible benchmarks that connect reasoning quality, collaboration efficiency, and human oversight within LLM-driven environments.
- *Absence of a holistic, Industry 5.0-oriented perspective:* Finally, most reviewed studies operate within an Industry 4.0 and 5.0 paradigm, emphasizing automation and optimization. Few frameworks explicitly align LLM-enabled manufacturing with the broader Industry 5.0 goals of resilience, sustainability, and human-centricity.

These gaps highlight the need for an approach that integrates cognitive intelligence with human oversight and cyber-physical grounding. In an attempt to address these deficiencies, we synthesized an exemplar conceptual structure, LLM-HSM, which systematically integrates large language models, human feedback mechanisms, digital twin systems, and verification/validation processes into a unified architecture. This initiative aims to operationalize adaptive learning, enhance resilience, and promote human-centric collaboration in next-generation manufacturing.

5. Exemplar conceptual structure of HITL and LLM integration

In this section, LLM is integrated with the already-developed H-SM framework [21], based on the literature and the identified gaps in the previous Section, to generate an exemplar and forward-looking structure, as shown in Fig. 10. It is built upon four key components: HITL, CPS, LLM, and VV&UM. The HITL module ensures seamless human-AI collaboration by providing intuitive user interfaces, workflow management, immersive environments, and decision-support systems, allowing operators to interact with and refine AI-driven processes. The CPS module bridges the gap between digital and physical manufacturing environments by integrating real-time data acquisition from IoT sensors and actuators with advanced DT simulations, enabling predictive analysis and adaptive process control. The LLM module serves as the AI-driven decision-making core, encompassing pre-training,

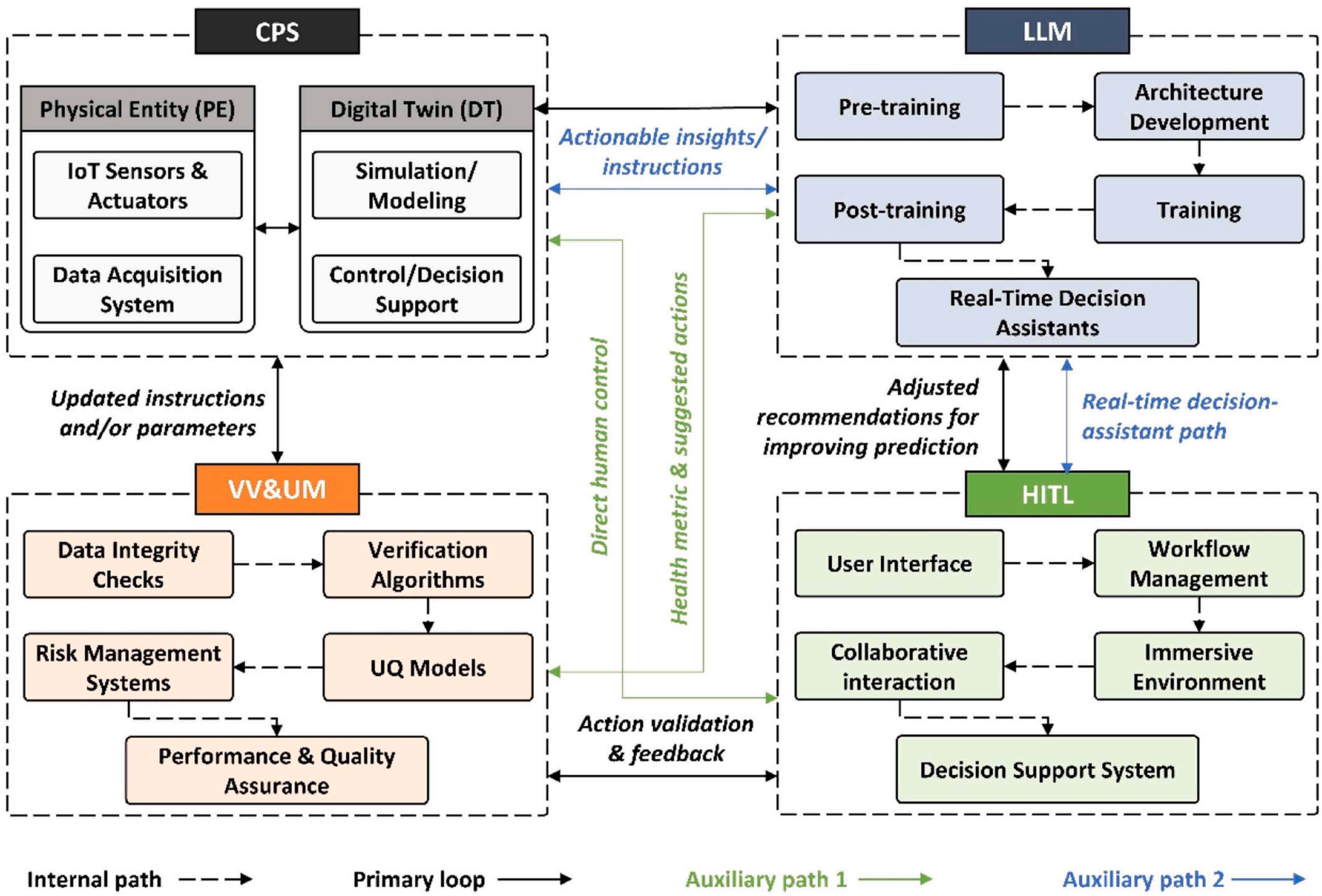

Fig. 10. Exemplar conceptual structure of large language model integrated with the human-in-the-loop-based smart manufacturing (LLM-HSM).

architecture development, post-training adaptation, and real-time decision assistants to optimize production, automation, and human-machine interaction. Finally, the VV&UM module ensures the safety, reliability, and accountability of AI-driven manufacturing decisions through data integrity checks, risk management systems, uncertainty quantification, and performance assurance mechanisms.

LLM-HSM can be situated within the broader paradigm of hybrid intelligence, which emphasizes the deliberate integration of human cognitive adaptability with machine-based computational capabilities to achieve outcomes neither could realize alone [204,205]. In this context, humans provide contextual reasoning, creativity, and ethical oversight, while LLMs and CPS infrastructures contribute scalability, speed, and data-driven insights. This symbiosis reflects a core principle of Industry 5.0, where human and artificial intelligences are designed to complement rather than compete.

5.1. Human-in-the-loop

The HITL module comprises four key components: user interface, workflow management, immersive environment, collaborative interaction, and decision support system. To facilitate cooperative decision-making and supervision in smart manufacturing processes, the HITL module serves as a crucial interface between human operators and the CPS [206]. The user interface component ensures smooth engagement by providing user-friendly tools, such as dashboards, augmented reality overlays, and multimodal inputs like speech and gesture detection. Thanks to this component, humans can effectively monitor and control automated processes, which improves operator situational awareness and streamlines complex operations [207]. Workflow management also streamlines resource allocation and task organization, guaranteeing seamless execution of operations for a variety of stakeholders. It reduces bottlenecks and inefficiencies by prioritizing processes and providing real-time updates through the integration of AI-driven insights.

To provide improved operator training, simulation, and real-time interactions with DTs, the HITL module also integrates an immersive environment, utilizing technologies such as AR and VR. By bridging the gap between virtual models and physical systems, these environments offer an interactive, hands-on experience that empowers operators to make well-informed decisions. This module is completed by the collaborative interaction and decision support system, which serves as a hub for collaboration between humans and machines. It guides users through challenging problem-solving scenarios, promotes collaboration in interdisciplinary settings, and offers actionable insights using sophisticated AI models and visualization tools. This comprehensive strategy ensures that human knowledge will always be integral to system functioning, thereby enhancing the robustness and adaptability of industrial workflows. This module addresses the gap of human exclusion in fully automated systems [208] by preserving adaptability, contextual reasoning, and ethical oversight [3]. This ensures that LLM outputs remain aligned with human judgment and organizational goals.

5.2. Cyber-physical system

The CPS module has two components: physical entity (PE) and DT, where the former consists of IoT sensors and actuators and a data acquisition system, and the latter consists of simulation/modeling and control/decision support. The CPS module forms the backbone of smart manufacturing, integrating physical and digital domains to achieve real-time process monitoring, simulation, and control. The PE component includes IoT sensors, actuators, and a robust data acquisition system that collectively enable the collection and transmission of real-world operational data. These sensors provide critical metrics such as temperature, pressure, and vibration, while actuators allow responsive adjustments to ensure system stability. The data acquisition system ensures that this information is accurately captured, filtered, and relayed to the DT for further analysis, bridging the gap between physical assets and their virtual counterparts [209].

The DT component is pivotal in mirroring the physical system and enabling predictive analysis, optimization, and decision-making. This component includes simulation/modeling tools replicating real-world processes to test scenarios, identify potential failures, and optimize operations before implementation. The control and decision-support subcomponent uses advanced algorithms to process real-time data and recommend actionable insights, ensuring system efficiency and reducing downtime [3]. Together, these components enable robust hardware and software integration, paving the way for real-time, adaptive, and intelligent manufacturing systems that can respond dynamically to evolving demands. The CPS module closes the gap between abstract LLM reasoning and physical reality by grounding decisions in sensor data, process models, and DT simulations. This prevents the detachment of LLM outputs from operational constraints, ensuring actionable and validated recommendations.

5.3. Large language model

LLMs form the core of the LLM-HSM structure, enabling advanced decision-making, automation, and human-machine interaction in smart manufacturing environments. At a high level, this section outlines the key stages in developing and deploying LLMs: pre-training, architecture development, training, post-training adaptation, and real-time decision assistance. Pre-training involves exposing the model to vast amounts of data to develop a foundational understanding of language. The architecture defines how the model processes information, leveraging scalable designs like the Transformer structure. Training refines the model's capabilities through specialized datasets and optimization techniques, while post-training focuses on fine-tuning the model for specific tasks or aligning its outputs with human values. Finally, the real-time decision assistant leverages these capabilities to provide actionable insights during manufacturing operations. The LLM module provides natural language interaction, knowledge integration, and contextual reasoning, directly addressing the challenge of accessibility and inclusivity in existing human-machine interfaces [210]. Below, we delve into each of these components in greater detail.

5.3.1. Pre-training

Compared with small-scale language models, LLMs require substantially larger volumes of high-quality data during pre-training, and their capacity and generalization ability are strongly shaped by both the composition of the pre-training corpus and the preprocessing strategies applied to it [65,211,212]. Contemporary LLMs are typically trained on heterogeneous textual data drawn from multiple sources, which can be broadly grouped into general-purpose data, such as webpages, books, and conversational text, and specialized data [67,213]. After data collection, rigorous preprocessing is essential to remove noise, redundancy, irrelevant content, and potentially harmful or sensitive information, as these factors have a direct impact on model robustness and downstream performance [214,215]. Recent efforts, such as Data Juicer, provide systematic toolchains for LLM-oriented data processing, offering extensive operators to support filtering, deduplication, privacy reduction, and tokenization [216]. A typical preprocessing workflow is illustrated in Fig. 11. Beyond preprocessing, effective data scheduling during pre-training is also critical, requiring careful control over both the proportion of different data sources in the training mixture and the order in which they are introduced, since these choices influence convergence behavior, knowledge retention, and the balance between general and specialized capabilities.

5.3.2. Architecture Development

The Transformer architecture, recognized for its parallel processing efficiency and scalability, has become the dominant foundation for LLM development, enabling models to scale to hundreds or even thousands of billions of parameters [59]. Modern LLM architectures can be broadly categorized into three types: encoder-decoder, causal decoder, and prefix decoder. The original Transformer adopts an encoder-decoder structure, where stacked self-attention layers in the encoder generate latent representations of the input sequence and the decoder applies cross-attention to autoregressively produce outputs, an approach that underpins pretrained language models such as T5 [217] and BART [218] and has been extended to LLMs like Flan-T5 [219]. In contrast, causal decoder architectures rely on unidirectional attention masks that restrict each token to attending only to preceding tokens, processing inputs and outputs within a single decoder stack, as exemplified by the GPT family, with GPT 3 demonstrating strong in-context learning capabilities at scale [61,64,211]. Prefix decoder architectures, also referred to as non-causal decoders [220], further modify this masking strategy by allowing bidirectional attention over prefix tokens [221] while preserving autoregressive generation for output tokens, thereby combining encoder-like contextual encoding with decoder-style generation using shared parameters. Rather than training from scratch, prefix decoders are often obtained by continuing training from causal decoders to accelerate convergence [222], as illustrated by U PaLM [223] derived from PaLM [65], with representative models including GLM 130B [224] and U PaLM [223].

5.3.3. Training

As model and data sizes continue to grow, training LLMs efficiently under limited computational resources has become increasingly challenging. Two main technical issues need to be addressed: enhancing training throughput and accommodating larger models within GPU memory. Several commonly used approaches to tackle these challenges include 3D parallelism [225,226] and mixed-precision training [227]. However, a detailed discussion of these methods is beyond the scope of this research.

The training of LLMs requires specific optimization techniques, scalable training methods, and memory-efficient approaches to improve efficiency and stability. Batch training plays a crucial role in pre-training, with state-of-the-art LLMs such as GPT-3 and PaLM utilizing dynamic batch size scaling from 32 K to 3.2 M tokens to enhance stability and throughput [65]. The learning rate follows a warm-up and decay strategy, which gradually increases in the early training phase and later decreases using a cosine decay schedule, optimizing convergence. Optimizers such as Adam [228] and AdamW [229] are commonly

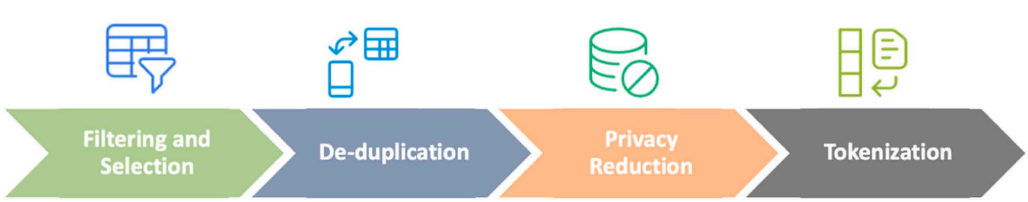

Fig. 11. Data preprocessing pipeline for pre-training LLMs.

employed, while Adafactor [230] is used in memory-constrained environments like PaLM and T5. As model sizes increase, scalable training techniques are required to manage computational resources efficiently. 3D parallelism, which combines data parallelism, pipeline parallelism, and tensor parallelism, is widely used to increase training throughput [231]. Data parallelism distributes training data across multiple GPUs, reducing memory constraints, while pipeline parallelism assigns consecutive model layers to different GPUs to optimize communication overhead [225,226]. Tensor parallelism further splits model weight matrices across GPUs, enabling large models to be trained efficiently while maintaining gradient synchronization. Open-source frameworks such as Megatron-LM, Colossal-AI, and DeepSpeed have implemented these techniques to scale training efficiently [232–234].

To further reduce computational costs, mixed precision training is employed, where 16-bit floating-point (FP16) and Brain floating-point (BF16) formats replace traditional 32-bit floating-point (FP32) calculations [227]. While FP16 reduces memory usage and increases computational efficiency on GPUs like NVIDIA A100, it may lead to precision loss, which BF16 mitigates by providing more exponent bits for improved representation accuracy [235]. Additionally, predictable scaling mechanisms have been introduced in GPT-4, allowing model performance forecasting with smaller models to optimize early-stage training [236]. In practice, multiple techniques are combined to maximize training efficiency. For example, BLOOM was trained using 8-way data parallelism, 4-way tensor parallelism, and 12-way pipeline parallelism across 384 NVIDIA A100 GPUs [235]. Open-source libraries such as DeepSpeed, PyTorch, and Alpa now provide built-in support for these techniques, including fully sharded data parallelism (FSDP) and ZeRO memory optimization, enabling the training of large models with minimal redundancy [237–239].

5.3.4. Post-training

After pre-training and training, LLMs develop generalized capabilities for solving a wide range of tasks. However, research has increasingly demonstrated that these capabilities can be further refined and adapted to meet specific objectives. Two primary approaches for adapting pre-trained LLMs are instruction and alignment tuning [10]. Instruction tuning focuses on enhancing or unlocking new abilities within the model, enabling it to perform specialized tasks more effectively. In contrast, alignment tuning ensures that the model's behavior aligns with human values, ethical considerations, or user preferences. Additionally, efficient tuning and quantization techniques are essential for optimizing LLMs in resource-constrained environments.

Instruction tuning is a fine-tuning approach that adapts pre-trained LLMs using a structured collection of natural language instructions [240]. This method is closely related to supervised fine-tuning [241] and multi-task prompted training [242]. The process begins with curating or constructing instruction-formatted instances, which are then used to fine-tune the LLM through supervised learning techniques, such as training with a sequence-to-sequence loss function. Studies have shown that after instruction tuning, LLMs exhibit enhanced generalization capabilities after instruction tuning, allowing them to effectively perform unseen tasks [219,240], including applications in multilingual environments [243]. A recent survey provides a comprehensive review of advancements in instruction tuning research, highlighting its impact on improving model adaptability and task performance [244].

5.3.5. Real-time decision assistant

The real-time decision assistant exploits the capabilities of LLMs to support immediate and context aware decision making in smart manufacturing environments by integrating real time data streams, expert knowledge, and adaptive learning mechanisms to assist human operators and automated systems, with the objective of improving operational efficiency, reducing errors, and enhancing overall system performance [118]. By processing heterogeneous inputs from sensor networks, manufacturing execution systems, human computer interfaces, and historical records, the LLM can analyze evolving conditions, identify correlations, and generate actionable insights through its natural language processing and long context handling capabilities, which is critical in manufacturing settings where system states change rapidly [245]. This component also strengthens human machine interaction through personalized and multimodal communication, enabling operators to query the system in natural language, receive context-aware guidance, and better understand complex processes [246], thereby supporting informed decision making [247].

Continuous feedback loops allow the LLM to refine its recommendations based on observed outcomes, ensuring ongoing improvement in relevance and accuracy [248]. Rather than operating in isolation, the real-time decision assistant is embedded within a broader smart manufacturing ecosystem and interoperates with complementary technologies, such as knowledge graphs, digital twins, and heuristic algorithms, to enhance interpretability, facilitate simulation-based evaluation of production scenarios, and support informed control decisions. Through this integration, routine cognitive tasks can be offloaded from human operators, reducing mental workload while maintaining human supervision, and in more advanced configurations, the LLM can autonomously execute specific functions such as design selection, process planning, equipment monitoring, CNC programming, and simulation-driven evaluation, thereby fostering an efficient and collaborative human-machine environment [118].

5.4. Verification, validation, and uncertainty management

The VV&UM module is crucial for ensuring the reliability, safety, and trustworthiness of an LLM-HSM system. It incorporates five key components: authentication and authorization algorithms, data integrity checks, uncertainty quantification models, risk management systems, and performance and quality assurance. As highlighted by Holzinger et al. [249], there is an urgent need for collaborative approaches to reevaluate oversight mechanisms in a time when AI could surpass human capabilities in comprehension. The goal is to create a robust system that can operate effectively in dynamic and complex manufacturing environments. This module mitigates core limitations of LLMs, such as hallucination, overconfidence, and lack of explainability, by embedding structured mechanisms for uncertainty estimation, verification against domain rules, self-reflection, and continuous validation through simulation and expert feedback [197,198,250]. This module incorporated the following components.

- *Authentication and Authorization Algorithms:* This component focuses on securing the system by ensuring only authorized users and devices can access it [251]. It uses robust algorithms to verify user identities and permissions, preventing unauthorized access and potential system misuse. Access control and encryption are implemented for data at rest and in transit to prevent data leaks. This ensures that sensitive data and processes are protected from both internal and external threats.
- *Data Integrity Checks:* Data integrity is paramount for reliable system operation; therefore, this component is designed to ensure that the data used in the system is accurate, consistent, and complete. Techniques such as data validation are used to maintain the quality of input data and detect any anomalies introduced by adversarial attacks. Additionally, quality filtering, data deduplication, and privacy reduction methods are employed to prepare the data for LLM training and implementation, thereby reducing data complexity and mitigating overfitting [54]. This involves conducting regular audits and monitoring the model's outputs to detect and address unintentional disclosures of sensitive information promptly.
- *Uncertainty Quantification Model:* This component addresses the inherent uncertainties in LLM predictions and manufacturing processes. It aims to quantify the level of uncertainty associated with LLM outputs, helping decision-makers assess the reliability of the

information provided by the system [248]. Methods include examining token probabilities and logits to determine output uncertainty, where a lower probability indicates a higher chance of hallucination [10]. Additionally, consistency checks involving multiple responses to the same question or using the LLM to reconstruct input questions from its response are used to assess the reliability of the model's outputs. The system can use external knowledge bases and contrastive learning techniques to mitigate uncertainty [75].

- *Risk Management Systems:* This component proactively identifies, evaluates, and mitigates potential risks associated with the operation of the LLM-based system. This includes adversarial testing, where ethical hackers attempt to exploit the model to identify vulnerabilities. The system monitors for data poisoning, adversarial prompting, and concept drift, all of which can degrade the model's performance. It incorporates mechanisms for handling model failures to maintain system safety and reliability [251].
- *Performance and Quality Assurance:* This component is responsible for continuously monitoring and evaluating the performance of the LLM and the overall system. It utilizes standardized evaluation metrics, such as BERT Score (precision, recall, F1-score), and others suitable for the specific tasks performed by the LLM. F1-micro and F1-macro scores are used to measure classification performance [251], while BLEU-4, ROUGE-L, and pass@10 are also employed in some cases to evaluate language generation [10]. This component oversees cross-validation, testing against real data, and employing feature mapping methods to ensure robustness and reliability. The continuous feedback loop from performance monitoring helps in adaptive learning and the ongoing refinement of the system [246]. This includes monitoring the model for biases and ensuring the model is not generating harmful or toxic content.

In summary, the VV&UM module is grounded in the established discipline of verification, validation, and uncertainty quantification (VVUQ), long applied in simulation-based engineering and cyber-physical systems [252]. By embedding structured validation, explainability, and confidence estimation into LLM-driven decision processes, the module directly addresses recognized risks such as hallucination, overconfidence, and lack of transparency, ensuring alignment with trustworthy AI principles in manufacturing [14].

5.5. Flow of information

LLM-HSM orchestrates information across four modules so that data, models, and human judgment reinforce one another. The primary loop begins in the PE, where IoT sensors and actuators generate real-time telemetry. A data acquisition layer filters and forwards this stream to the DT for state estimation and predictive simulation. DT outputs, current operating state, constraint checks, and what-if forecasts are then provided to the LLM, which fuses these signals with domain knowledge to produce context-aware recommendations. These recommendations are presented through HITL interfaces (dashboards, AR overlays, decision support), where operators validate or adjust them. Approved actions are issued back to the PE for execution, closing the loop.

Two auxiliary paths strengthen robustness and responsiveness. First, direct human control via HITL allows operators to monitor or intervene in CPS without traversing the LLM when urgent manual actions are required (e.g., safety stops or parameter overrides). Second, a real-time decision-assistant path enables operators to query the LLM on demand (explanations, alternatives, reports) while maintaining traceability to the DT state and plant constraints. Throughout all loops and paths, the VV&UM module operates continuously, authenticating access, verifying data integrity, quantifying uncertainty in LLM outputs, and enforcing validation against domain rules and DT simulations before recommendations are enacted. This integrated flow ensures recommendations are grounded (CPS data), auditable (VV&UM checks and uncertainty exposure), and human-aligned (HITL oversight). By coupling fast, data-driven inference with expert judgment and explicit validation, the structure avoids over-automation and reduces the risk of ungrounded or overconfident outputs, thereby delivering clear and reproducible information pathways.

5.6. Integration rationale

Each module in the LLM-HSM structure addresses different gaps in current manufacturing systems, but their true potential emerges when they operate in concert. The complementarities between these modules are essential for bridging limitations observed in prior work. Studies on digital twins in manufacturing, for example, demonstrate their effectiveness in process monitoring and predictive control, but also highlight difficulties in adapting models to rapidly changing conditions without human oversight [253]. Similarly, research on LLM-enabled digital twins shows promise in automating tasks such as fault detection or scheduling, yet acknowledges vulnerabilities to hallucinations and overconfidence when applied in safety-critical contexts [254]. By integrating HITL, we ensure that such limitations are counterbalanced by human reasoning, while VV&UM provides transparency and uncertainty estimation to prevent over-reliance on machine outputs. This interplay among modules prevents the system from becoming either overly automated or excessively dependent on fallible human judgment.

A further rationale for integration lies in mitigating typical failure modes of LLMs within manufacturing environments. Hallucination and lack of grounding are countered through DT simulation and sensor-based validation, which constrain generative outputs within feasible process boundaries. Overconfidence and limited explainability are mitigated through VV&UM modules that embed confidence scoring, error tracking, and structured validation pipelines, echoing recent calls for trustworthy AI in industrial applications [14]. The risks of operator exclusion or steep learning curves are reduced by natural language interfaces provided by LLMs, while HITL ensures that inclusivity and ethical oversight remain central. In this way, integration is not a design choice of convenience but a necessity to ensure that the weaknesses of one module are consistently compensated by the strengths of another, yielding a reliable, adaptive, and human-aligned system for next-generation manufacturing.

5.7. Mapping manufacturing KPIs to the modules

While Section 2.2.3 reviews manufacturing-specific key performance indicators reported in recent LLM-enabled manufacturing studies, their role within the proposed LLM-HSM structure requires explicit clarification. As an exemplar conceptual structure, LLM-HSM does not claim direct numerical optimization of KPIs; instead, it defines the functional responsibilities through which different system modules influence performance outcomes. Mapping KPIs to these modules clarifies how human centricity, cyber-physical integration, language-based reasoning, and reliability assurance jointly contribute to measurable manufacturing objectives. This mapping also strengthens the connection between the background review and the conceptual contribution by illustrating how reviewed KPIs can be operationally interpreted within the proposed framework. Table 4 summarizes representative manufacturing KPIs and associates them with the LLM HSM modules that primarily govern or influence them, together with the underlying mechanisms of influence.

6. Features and technologies

6.1. Identification of key enabling features and technologies

The successful implementation of the LLM and HITL concepts in smart manufacturing relies on several key enabling features and technologies that ensure seamless human-AI collaboration, real-time

Table 4
Mapping between LLM-HSM modules and representative manufacturing KPIs.

Modules	Representative manufacturing KPIs	Primary mechanism of influence
HITL	Human error rate Decision accuracy Time to competence Operator workload Safety incidents	– Human oversight/ contextual judgment for AI-generated recommendations – Enables corrective intervention in critical decisions – Supports learning and skill acquisition through interactive assistance – Mitigates cognitive overload and unsafe autonomous behavior
CPS	Cycle time Makespan Overall equipment effectiveness Mean time to detection Mean time to repair	– Real-time monitoring/synchronization between physical & digital assets – Supports simulation-based evaluation and predictive maintenance – Supplies reliable operational data for decision support
LLM	Planning latency Setup time Design iteration count Schedule adherence Response time to disturbances	– Facilitates reasoning over heterogeneous data sources – Generates and interprets natural language instructions – Accelerates planning and reconfiguration tasks – Supports adaptive decision making across the manufacturing lifecycle
VV&UM	False alarm rate Validation coverage Confidence level of recommendations Risk exposure Compliance indicators	– Governs verification and validation of AI outputs – Performs uncertainty estimation and risk assessment – Ensures reliability, traceability, and trustworthiness of LLM-driven decisions in safety-critical environments

decision-making, and process optimization. They bridge the gap between advanced AI capabilities and practical manufacturing constraints, allowing LLMs to function effectively in dynamic industrial environments. This integration enhances flexibility, reliability, and operational efficiency by integrating intelligent automation, adaptive CPS, HITL interaction, and rigorous verification mechanisms. This section examines the fundamental components that enable LLM and HITL, including interactive human-machine interfaces, real-time DT integration, LLM-driven contextual decision support, and robust verification and uncertainty management techniques. These features and technologies collectively enable a smarter, safer, and more adaptable manufacturing ecosystem, ensuring that AI-driven innovations align with industry requirements and human expertise.

Key-enabling features and technologies were identified based on their frequency of appearance in reviewed papers extracted using the methodology outlined in Section 2, as well as relevant studies and discussions among the co-authors. They were categorized based on their applications in each module. Fig. 12 schematically presents to which module each feature and technology belongs. Some of them are common between two or even three modules, while some are only in correspondence with one of the modules. For instance, Knowledge Graph (KG) is used in both LLM and VV&UM modules, while High-performance computers are employed for all modules. This is presented by separated ribbons in Fig. 12. In addition, for a common understanding, these terms are redefined based on the literature to establish a common ground for future research.

Table 5 lists the definitions and examples of the thirty-four key-enabling features and technologies.

6.2. Synergies and trade-offs

The key enabling technologies identified in Section 6.1 do not function as independent add-ons; their value arises from their composition within LLM-HSM, which supports HITL decision-making. In the proposed architecture, the LLM does not operate as a monolithic black-box assistant but rather as a cognitive hub that interfaces with all

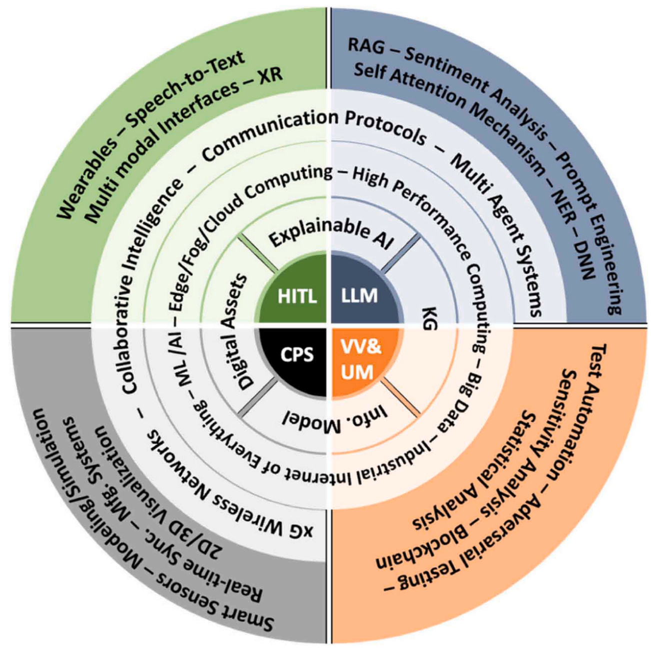

Fig. 12. Features and technologies in LLM-HSM.

modules. This composition yields several capabilities that existing frameworks do not combine systematically. First, it enables the system to generate explanations and recommendations that are both natural language and structurally consistent with plant semantics (e.g., machine-product-operation relationships), rather than free-text suggestions detached from the actual asset model. Second, it enables division of labor between agents specialized for scheduling, code synthesis, maintenance, or training tasks, while the HITL and VV&UM modules retain global oversight and safety constraints. This differs from classical CPS architectures, where optimization and diagnosis modules are usually task-specific and cannot flexibly reconfigure their roles in response to new instructions or organizational changes. Third, by embedding digital twins into the loop as simulation and verification back-ends, LLM-generated plans and control suggestions can be stress-tested and refined before deployment, enabling a form of self-correcting behavior [114] that is not present in static knowledge-based or purely data-driven pipelines.

At the same time, LLM-HSM makes explicit trade-offs between autonomy, coordination, fidelity, and responsiveness. Increasing agent autonomy (e.g., allowing scheduling or maintenance agents to execute actions semi-independently) can improve responsiveness and reduce human workload, but it also raises the risk of local decisions that conflict with global production objectives or safety policies. Integrating HITL and LLM with smart manufacturing through structures such as LLM-HSM addresses this by routing critical actions through the VV&UM and HITL modules, which act as coordination and override points rather than simple monitoring components. Similarly, enriching the knowledge graph with fine-grained process, asset, and safety constraints improves the fidelity and interpretability of LLM reasoning, but also increases the cost of maintaining the graph and can introduce latency in real-time scenarios. LLM-HSM explicitly separates slow-changing structural knowledge (e.g., process routes, capabilities, safety rules) from fast-changing operational data (e.g., sensor readings, work-in-progress (WIP)), delegating the former to the knowledge graph and the latter to DT/CPS streams to balance fidelity and responsiveness. A further trade-off arises between generative flexibility and verification cost: more expressive LLM prompts and agent roles can cover wider task spaces, but they also require more extensive testing in the digital twin environment; LLM-HSM addresses this by defining risk-dependent verification regimes, where high-risk recommendations are always simulated or

Table 5
Key enabling features and technologies in LLM-HSM, their definitions, and their relevance to HITL, CPS, LLM, and VV&UM.

Key-enabling features and technologies	Relevant module - Definition	Ref.
Wearables	HITL - Wearable devices are electronic devices that can be worn on the body, often designed to track and monitor various types of data about the user and their environment	[165]
Speech-to-Text (STT)	HITL - STT, also known as automatic speech recognition (ASR), is a technology that converts spoken language into written text. This process is a crucial component of many human-computer interaction systems, particularly those involving conversational agents	[255]
Multi-modal Interfaces	HITL - Multi-modal interfaces are systems that can process and respond to a variety of input types, or modalities, such as text, speech, images, and sensor data. These interfaces enhance the richness and flexibility of human-computer interaction, allowing for more intuitive and comprehensive communication	[43, 118, 147]
Extended Reality (XR)	HITL - XR is a broad term that encompasses immersive technologies, including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). These technologies aim to blend or replicate the real world with a "DT reality" that can interact with it.	[255, 256]
Smart Sensors/ Actuators	CPS - Apart from their primary use, smart sensors also provide the interfaces required for integration into IIoT networks. These could be XR-based gadgets or wireless networks.	[257]
Modeling & Simulation	CPS - modeling involves creating a representation of a system or process, while simulation involves using that representation to imitate the behavior of the system over time. Models can be used as a basis for simulation. Models can be physics-based, physics-driven, or physics-informed.	[258]
Real-time Synchronization	CPS - Real-time synchronization in manufacturing refers to the coordination of processes and resources by providing the right components to subsequent production steps at the right time. It is closely related to the Just-in-Time (JIT) philosophy, aiming for integrated and synergistic production organization and management.	[259]
Manufacturing Systems	CPS - It can be defined as a combination of humans, machinery, and equipment that are bound by information flow with the aim of establishing a smart factory.	[260]
2D/3D Visualization	CPS - 2D visualizations present information on a flat plane, using elements like charts, graphs, and images. These are commonly used for displaying data in a structured and easily digestible manner. 3D visualizations represent data or objects in a three-dimensional space, providing a more realistic and immersive view	[261]
Sentiment Analysis	LLM - Sentiment analysis is a process that analyzes text to determine the emotional tone or attitude expressed within it. It can be used to understand the opinions, feelings, and evaluations of individuals or groups regarding a particular topic, product, or service.	[255]

Table 5 (*continued*)

Key-enabling features and technologies	Relevant module - Definition	Ref.
Prompt Engineering (PT)	LLM - PT is the process of designing effective prompts to elicit desired responses from LLMs. It involves crafting specific instructions or questions that guide the LLM to generate the intended output	[262]
Named Entity Recognition (NER)	LLM - NER is an NLP task that involves recognizing and categorizing specific entities within a text. These entities may include proper nouns or terms that refer to real-world objects or concepts.	[145]
Deep Neural Network (DNN)	LLM - DNN is a type of artificial neural network (ANN) that uses multiple layers to analyze data. It is a core component of deep learning (DL) and is designed to mimic how the human brain processes information.	[263]
Retrieval-Augmented Generation (RAG)	LLM/ VV&UM - RAG is a hybrid AI approach that merges information retrieval with generative language models to generate more precise and contextually relevant responses. It enhances LLMs by incorporating dynamic, external knowledge sources.	[155]
Self-Attention Mechanism	LLM - The self-attention mechanism is a key element of the Transformer architecture, commonly used in LLMs. It enables the model to assess the relevance of different parts of the input sequence during processing, allowing it to capture long-range dependencies and relationships between tokens.	[118]
Test Automation	VV&UM - Test automation refers to the use of software tools and scripts to perform testing activities that are usually done manually. It is a crucial part of software development and quality control, particularly in manufacturing and industrial automation where it is essential to ensure reliability, safety, and efficiency.	[258]
Blockchain	VV&UM - It is a distributed system with growing lists of blocks that are securely linked together and can be used for different purposes like authentication, smart contracts, and security protocols.	[264]
Adversarial Testing	VV&UM - Adversarial testing is a method used to evaluate the robustness and reliability of AI systems, particularly machine learning models, by intentionally trying to "break" them. This is achieved by exposing the AI to carefully crafted inputs designed to elicit incorrect or unexpected behavior. The goal is to identify the model's weaknesses and vulnerabilities, which can then be addressed to make the system more dependable.	[45, 263]
Sensitivity Analysis (SA)	VV&UM - SA is a method used to understand how changes in the input of a model affect its output. It is a crucial part of model development, validation, and decision-making, especially in complex systems like industrial automation.	[265]
Statistical Analysis	VV&UM - Statistical analysis is a broad term referring to the collection, analysis, interpretation, presentation, and organization of data using various statistical methods and techniques. It is used to identify patterns, trends, and relationships in data, and to draw meaningful conclusions, which can inform decision-making in a variety of fields.	[179]

(*continued on next page*)

Table 5 (*continued*)

Key-enabling features and technologies	Relevant module - Definition	Ref.
Digital Asset	HITL/ CPS - It is everything that is exclusively available digitally and has a clear usage right or license. Digital assets include but are not exclusive to digital documents, audible content, motion pictures, non-fungible tokens (NFTs), and other relevant digital data.	[266]
Explainable AI (XAI)	HITL/ LLM - XAI refers to AI systems designed for transparency and interpretability, enabling humans to comprehend how decisions or predictions are made. It is a research area focused on making AI models' inner workings more understandable to people.	[263]
Information Model	CPS/ VV&UM - An information model is a way of structuring raw data into a logical, hierarchical structure, enriched with contextual meta-descriptions. Information models are designed for machine understanding. They transform data into structured information	[267]
Knowledge Graph (KG)	LLM/ VV&UM - A KG is a structured representation of knowledge that uses a graph structure to model and store information. It is a key component of cognitive intelligence. KGs represent knowledge as a network of interconnected entities and their relationships, offering a way to organize and access information	[194, 268]
Multi Agent System (MAS)	HITL/ CPS/ LLM - MAS is a system made up of multiple intelligent agents that interact to accomplish individual or collective objectives. These agents can be software, robots, or humans, and they function within a common environment.	[269]
Next-generation (NextG, xG) wireless Networks	HITL/ CPS/ LLM - xG wireless networks, like 5 G and 6 G, are advanced communication systems developed to address growing needs for faster speeds, better connectivity, and enhanced reliability in a connected world. These networks focus on boosting data transfer rates, minimizing latency, expanding network capacity, and ensuring secure and dependable connectivity.	[270]
Collaborative Intelligence	HITL/ CPS/ LLM - Collaborative intelligence refers to the synergistic combination of human and AI to achieve outcomes that neither could accomplish independently.	[118, 271]
Communication Protocols	HITL/ CPS/ LLM - Communication protocols are standardized sets of rules that enable different devices or systems to exchange information in a structured and consistent way. They are essential for ensuring interoperability and reliable communication in complex systems, such as those found in manufacturing environments.	[267, 272]
Internet of Everything (IoE, IoX)	HITL/ CPS/ LLM/ VV&UM - IoE is an evolution of the IoT that integrates devices, processes, data, and people into a single, highly interconnected environment. The IoE aims to put the experiences and needs of human beings at the center of industrial ecosystems. It involves not only smart devices but also humans, robots, and processes within the network.	[273]
Machine Learning (ML)/ Artificial Intelligence (AI)	HITL/ CPS/ LLM/ VV&UM - ML and AI are interconnected but different concepts. AI is a wide-ranging term that includes various methods for enabling systems to perform tasks usually requiring human intelligence, whereas ML is a subset of AI that focuses on allowing systems to learn from data.	[265]
Edge/ Fog/ Cloud Computing	HITL/ CPS/ LLM/ VV&UM - Location is the primary distinction between cloud, fog, and edge computing. While fog and edge are decentralized resources where data is processed closer to edge devices, cloud computing is a centralized approach where data is stored, processed, and retrieved from a distant data center.	[274, 275]
High-Performance Computing (HPC)	HITL/ CPS/ LLM/ VV&UM - HPC involves the processing of large volumes of data and the handling of substantial model sizes. This requires significant computational resources that are beyond the capabilities of standard desktop computers	[276]
Big Data	HITL/ CPS/ LLM/ VV&UM - Big data refers to the extremely large and complex datasets that are generated in modern manufacturing and other industrial settings. These datasets are often characterized by their volume, variety, and velocity, and they pose significant challenges for traditional data processing and analysis techniques	[82]

sandboxed before execution, while low-risk suggestions can be applied with lighter checks.

7. Discussion

7.1. Relevance to industry 5.0

Integration of HITL and LLM, which is demonstrated through the exemplar LLM-HSM structure in this paper, aligns strongly with the overarching goals of Industry 5.0, resiliency, sustainability, and human-centricity [277], by leveraging HITL intelligence GenAI to create adaptive, knowledge-driven, and collaborative manufacturing systems. This enhances manufacturing resilience through its dynamic, human-AI feedback loops and multimodal learning capabilities. By continuously integrating real-time data from DTs, sensors, and production environments, the structure enables predictive responses to disturbances, anomalies, or operational shifts. The LLM component acts as a cognitive layer that interprets data patterns, generates adaptive control suggestions, and supports decision-making under uncertainty [278]. Simultaneously, human operators provide contextual judgment and domain-specific insight, ensuring that the system remains robust against unforeseen disruptions, anomalies, or supply chain fluctuations.

In addition, sustainability from optimization-oriented decision support and knowledge reusability mechanisms. LLM-HSM utilizes LLM-based reasoning to minimize resource consumption, energy usage, and waste generation by identifying more efficient production pathways and developing adaptive scheduling strategies. Integration with DTs allows continuous evaluation of energy profiles, material flow, and lifecycle performance, facilitating real-time adjustments that reduce environmental impact. Moreover, HITL collaboration ensures that sustainability considerations, such as ethical sourcing, recycling strategies, and green manufacturing, are incorporated into generative planning and simulation processes [278]. This synergy between cognitive automation and human ethical oversight contributes to sustainable production aligned with Industry 5.0's environmental and social objectives.

And finally, at its core, LLM-HSM embodies the human-centric concept of Industry 5.0 by positioning humans as central orchestrators rather than peripheral operators. Through natural-language interfaces and contextual understanding, LLMs enable intuitive human-machine interaction and lower cognitive barriers. Humans are empowered to guide, validate, and refine model-generated insights, fostering trust, transparency, and explainability [279]. The structure thus facilitates a symbiotic partnership in which human creativity, ethics, and experiential knowledge complement AI-driven reasoning and automation.

7.2. Socio-technical and organizational implications

While integration of LLM and HITL into manufacturing advances technical resilience, its implications in real industrial environments are shaped equally by socio-technical and organizational factors. Successful deployment depends not only on cognitive automation but also on institutional readiness, evolving workforce roles, and governance structures that ensure responsible and transparent use of generative AI [280,281]. From an organizational learning perspective, this integrative forward-looking paradigm shifts manufacturing knowledge from static documentation and tacit operator experience toward continuously updated, conversational knowledge repositories. This requires mechanisms for validating, curating, and institutionalizing model-generated insights. Plants must therefore invest in processes for human-AI co-learning, including feedback channels where operators correct hallucinations, refine task guidelines, and gradually encode their expertise into the system. Such feedback loops support collective learning rather than one-off automation.

In terms of workforce transformation, exemplary structures such as LLM-HSM accelerate the transition from manual machine operation toward supervisory, interpretive, and diagnostic roles. Operators interact with multimodal LLMs through natural language, reducing technical barriers but increasing the need for digital literacy, critical AI judgement, and socio-technical awareness. Reskilling and continuous training thus become essential complements to the architecture, ensuring that workers can understand, challenge, and govern AI-driven recommendations rather than merely follow them. In addition, ethical and governance considerations are also integral. Because generative models can introduce biased suggestions, opaque reasoning, or unsafe control actions, factories adopting updated structures such as LLM-HSM must establish clear policies for auditability, traceability, and human override authority. Formal guidelines for data provenance, model updates, and decision accountability are needed to prevent over-automation and ensure alignment with regulatory expectations for safety-critical systems. This is particularly relevant where LLMs influence scheduling, dynamic control, or maintenance decisions that directly affect operator safety.

Finally, policy alignment frames how revolutionized approaches such as LLM-HSM scale beyond individual case studies. Industry 5.0 initiatives are increasingly emphasizing trustworthy AI and human empowerment [14], the green transition [282], and cybersecurity [283]. We aim to contribute to these agendas only when embedded in a broader ecosystem of standards, worker protections, certification procedures, and European AI regulatory frameworks. Its success thus depends on coordinated socio-technical adoption rather than purely technical maturity.

7.3. Ethical and human rights safeguards

Within LLM-HSM, ethical and human-rights considerations are not treated as abstract principles but are operationalized through explicit technical and organizational mechanisms [14]. First, explainability is embedded into the interaction between the LLM, HITL, and VV&UM modules. For high-impact recommendations, such as changes in process parameters, rescheduling decisions, or maintenance deferrals, the LLM is required to generate a human-readable rationale that references the relevant data sources, constraints, and rules involved. These rationales are presented to operators and supervisors, along with risk indicators, enabling informed judgment rather than blind acceptance of model outputs [284,285]. This supports transparency and helps prevent opaque or arbitrarily biased decisions in safety-critical contexts.

Second, auditability is ensured through systematic logging and traceability. Prompts, retrieved knowledge-graph entities, model versions, and the sequence of actions proposed and executed are stored in an auditable record that can be inspected during incident analysis, compliance checks, or post-hoc reviews. This logging infrastructure enables organizations to reconstruct how a recommendation emerged, which constraints were applied, and which human approvals were involved, thereby supporting accountability and alignment with emerging regulatory requirements for high-risk AI systems [286]. Additionally, access control and role-based views restrict who can query or modify specific models, datasets, and decision pipelines, thereby protecting sensitive operational and worker-related information.

Third, LLM-HSM formalizes human control and worker protection by design. Decisions that may affect worker safety, workload, or employment conditions are always routed through the HITL and VV&UM modules, which function as mandatory approval gates rather than passive monitoring components. Operators can override, delay, or request clarification for any LLM-generated suggestion, and their feedback is used to refine prompts, constraints, and verification regimes over time. This preserves human agency, supports non-discrimination and dignity at work, and aligns the use of generative AI with the human-centric and rights-oriented vision of Industry 5.0.

From a structural perspective, these ethical safeguards are structurally embedded in the core modules of LLM-HSM. The HITL module operationalizes human agency by ensuring that high-impact decisions remain subject to human review, interpretation, and override. Complementarily, the VV&UM module provides a technical control layer that evaluates, verifies, and filters LLM-generated outputs before they can influence operational processes. Together with the CPS and LLM components, these modules create a governance mechanism that integrates explainability, accountability, and human supervision directly into the system's decision pipeline rather than applying them only at the organizational level.

7.4. Challenges

Generative artificial intelligence has great potential to improve industrial practices and design, yet a close examination of current LLM capabilities across multiple benchmarks reveals significant limitations. One pressing concern lies in reasoning: LLMs often underperform in analytical and spatial reasoning tasks, which are critical in design, simulation, and manufacturing. This limits their ability to execute complex computations or geometric problem-solving, even when domain-specific prompts are used. A further challenge is the risk of human oversight errors and cognitive overload. Although HITL interaction increases reliability, it may also lead to operator fatigue, decision paralysis, or a reliance on model outputs without critical evaluation. Inexperienced operators in particular may accept erroneous recommendations due to a lack of domain expertise or misplaced confidence. Effective interface and alert design must therefore carefully balance interpretability and cognitive demand.

Bias and domain generalization also remain persistent obstacles. Since LLMs are typically trained on general-purpose data, their outputs often fail to capture the semantics or procedures of manufacturing environments. Even fine-tuned models may carry blind spots due to the scarcity of annotated industrial datasets. Compounding this issue is the problem of hallucination: confident but incorrect outputs, which in safety-critical settings could cause severe consequences. While techniques such as retrieval-augmented generation and confidence scoring can mitigate hallucination, their industrial deployment remains experimental. Additionally, environmental and scalability issues pose further barriers. Training and deploying large-scale models consume vast energy resources, raising sustainability concerns, especially for small- and medium-sized enterprises. Moreover, when tasked with increasingly complex or multi-step processes, current LLMs show degraded performance, underscoring the need for modular or incremental task structuring.

Socioeconomic and legal challenges also deserve attention. The rise of AI copilots may displace workers in design, modeling, and documentation roles, requiring proactive investment in reskilling. Intellectual property questions also remain unresolved: LLMs trained on large datasets risk inadvertently reproducing elements of patented designs, raising concerns about originality, authorship, and potential legal liability. Finally, privacy and security risks surrounding sensitive design and manufacturing data can lead to financial losses or competitive disadvantages if compromised. Overall, these challenges, ranging from technical and environmental to social, legal, and ethical, underscore the urgent need for trustworthy, transparent, and human-centered frameworks before LLMs can be reliably integrated into manufacturing systems.

7.5. Operationalization pathway of the exemplar LLM-HSM structure

The integration of LLMs and HITLs into manufacturing paradigms and Industry 5.0, through approaches such as LLM-HSM, is conceived as a process to be instantiated incrementally rather than deployed as an all-or-nothing solution. A pragmatic operationalization pathway is to introduce LLM-HSM in stages, each building on existing CPS infrastructure and progressively tightening the human-AI-DT loop. Fig. 13 illustrates this progressive operationalization pathway.

In an initial stage, the LLM layer operates as an offline advisory system, consuming historical and live CPS data to support tasks such as anomaly explanation, log summarization, parameter suggestion, and documentation generation, while all control actions remain executed through conventional automation and human procedures. This enables early assessment of usefulness and trust without altering safety-critical loops.

In a second stage, the LLM is coupled more tightly with the knowledge graph and digital twin to enable grounded, simulation-backed recommendations. Here, the VV&UM and HITL modules explicitly determine which LLM proposals are tested, accepted, or rejected. For example, proposed schedule changes, parameter adjustments, or maintenance plans are first translated into structured actions, evaluated in the digital twin, and only then presented to operators together with explanations and risk indicators. In a third stage, selected LLM-driven agents may be allowed to trigger semi-automated actions under clearly defined risk thresholds and override policies, for instance by issuing low-risk set-point changes or generating code modifications that are automatically verified in a sandbox environment before deployment. At each stage, the effect of LLM-HSM should be evaluated against baseline CPS setups using manufacturing KPIs such as scrap and rework rates, time-to-diagnose anomalies, programming and setup effort, downtime, and operator workload. This staged pathway provides a concrete route for transforming LLM-HSM from a conceptual architecture into an operational, measurable augmentation of existing smart manufacturing systems.

For this operationalization pathway, small and medium-sized enterprises (SMEs) are particularly critical. SMEs represent the majority of manufacturing firms and often operate with constrained engineering resources, fragmented IT infrastructures, and limited in-house AI expertise, yet they face the same pressures for flexibility, customization, and sustainability as large enterprises. The modular, staged nature of LLM-HSM directly targets these conditions by allowing SMEs to begin with low-risk, advisory use cases that leverage existing machines, logs, and basic CPS components, and only later progress towards tightly coupled, simulation-backed, and semi-automated configurations as capabilities and trust mature. In this sense, SMEs are not only beneficiaries but also key testbeds: if the framework can be made deployable, maintainable, and interpretable under SME constraints, it is more likely to be scalable and robust across the broader manufacturing ecosystem.

8. Outlook

The exemplar LLM-HSM conceptual structure synthesized in this study lays the foundation for a modular, human-centered integration of LLMs, CPS, HITL, and VV&UM. While this paper emphasizes core manufacturing operations, future research should extend its scope to adjacent domains such as supply chain management, testing and validation, customer service, and end-of-life processes like recycling and reuse, where LLMs already show emerging potential. We therefore position LLM-HSM as a conceptual yet necessary perspective on how to integrate LLMs, HITL decision-making, and CPS infrastructures within an Industry 5.0 vision. We acknowledge that this must be empirically validated and needs a proof of concept (PoC), and our ongoing work aims to instantiate and assess LLM-HSM in a wire arc additive manufacturing (WAAM) setting, where an LLM layer will be integrated with process monitoring, parameter recommendation, and code review loops around an existing CPS stack. In this context, we plan to compare baseline WAAM operation against an LLM-HSM enhanced configuration using KPIs such as defect incidence, rework and tuning effort, and responsiveness to process disturbances, providing an initial systematic test of the framework's practical value.

Beyond these implementation efforts, theoretical extensions are

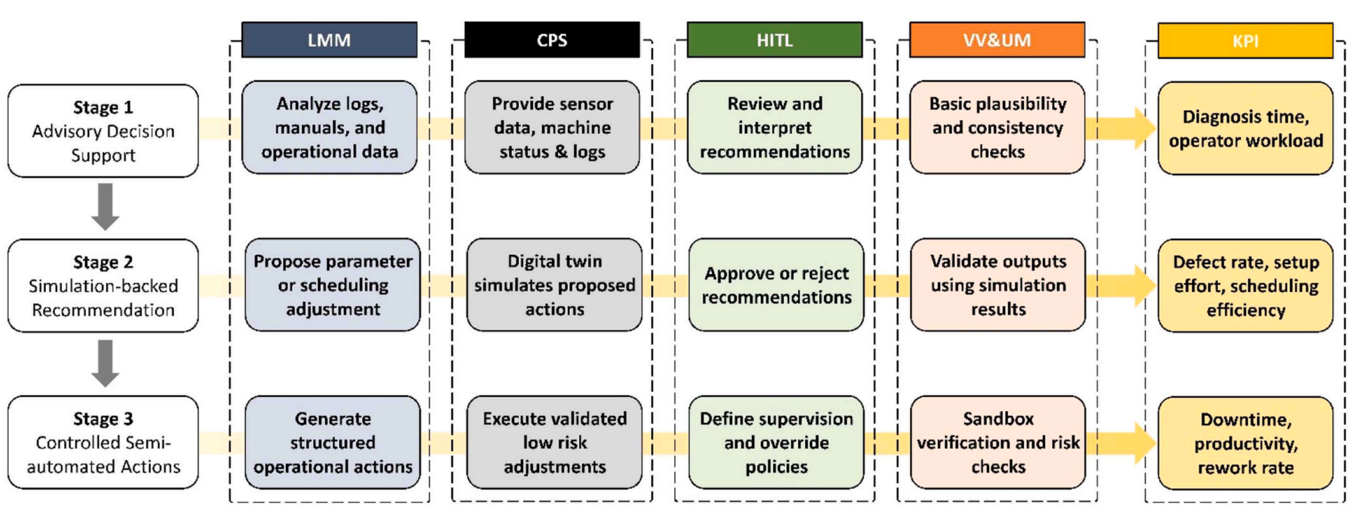

Fig. 13. Staged operationalization pathway of LLM-HSM.

needed to systematically evaluate human-AI collaboration. Future work should define metrics such as an adaptability index to quantify co-adaptation in dynamic settings, a trust calibration score to measure the alignment between operator confidence and model reliability, and uncertainty-aware KPIs that assess robustness, explainability, and performance simultaneously. Incorporating cognitive workload and operator satisfaction into these evaluation frameworks would provide a more holistic measure of hybrid intelligence in Industry 5.0. By pursuing these directions, the field can move from conceptual architectures to operational systems that embody the human-centric, resilient, and sustainable values of Industry 5.0.

CRediT authorship contribution statement

Duhwan Mun: Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Duck Bong Kim:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Mahdi Sadeqi Bajestani:** Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization.

Declaration of Competing Interest

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

Acknowledgement

This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the Global Research Support Program in the Digital Field program (RS-2024–00411653) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and by the Basic Science Research Program (NRF-2022R1A2C2005879) through the National Research Foundation of Korea (NRF) funded by the MSIT, Korea. Special thanks to Mohammad Mahruf Mahdi, whose comments and suggestions improved the manuscript.

References

[1] Phuyal S, Bista D, Bista R. Challenges, opportunities and future directions of smart manufacturing: a state of art review. Sustain Futures 2020;2. https://doi.org/10.1016/j.sftr.2020.100023.
[2] Wang B, Tao F, Fang X, Liu C, Liu Y, Freiheit T. Smart manufacturing and intelligent manufacturing: a comparative review. Engineering 2021;7. https://doi.org/10.1016/j.eng.2020.07.017.
[3] Wang B, Zheng P, Yin Y, Shih A, Wang L. Toward human-centric smart manufacturing: a human-cyber-physical systems (HCPS) perspective. J Manuf Syst 2022;63. https://doi.org/10.1016/j.jmsy.2022.05.005.
[4] Filippi E, Bannò M, Trento S. Automation technologies and their impact on employment: a review, synthesis and future research agenda. Technol Forecast Soc Change 2023;191. https://doi.org/10.1016/j.techfore.2023.122448.
[5] Thomas D, Weiss B. Maintenance costs and advanced maintenance techniques in manufacturing machinery: survey and analysis. Int J Progn Health Manag 2021:12. https://doi.org/10.36001/ijphm.2021.v12i1.2883.
[6] Evjemo LD, Gjerstad T, Grøtli EI, Sziebig G. Trends in smart manufacturing: role of humans and industrial robots in smart factories. Curr Robot Rep 2020;1. https://doi.org/10.1007/s43154-020-00006-5.
[7] Ghobakhloo M, Mahdiraji HA, Iranmanesh M, Jafari-Sadeghi V. From industry 4.0 digital manufacturing to industry 5.0 digital society: a roadmap toward human-centric, sustainable, and resilient production. Inf Syst Front 2024. https://doi.org/10.1007/s10796-024-10476-z.
[8] Negri M, Cagno E, Colicchia C, Sarkis J. Integrating sustainability and resilience in the supply chain: a systematic literature review and a research agenda. Bus Strategy Environ 2021;30. https://doi.org/10.1002/bse.2776.
[9] Zhou J, Zhou Y, Wang B, Zang J. Human–cyber–physical systems (HCPSs) in the context of new-generation intelligent manufacturing. Engineering 2019;5. https://doi.org/10.1016/j.eng.2019.07.015.
[10] Zhao W.X., Zhou K., Li J., Tang T., Wang X., Hou Y., A survey of large language models Cornell University arXiv2023. 10.48550/arXiv.2303.18223.
[11] Balasubramanian V. Brain power. Proc Natl Acad Sci - PNAS 2021;118. https://doi.org/10.1073/pnas.2107022118.
[12] Gebicke-Haerter PJ. The computational power of the human brain. Front Cell Neurosci 2023;17. https://doi.org/10.3389/fncel.2023.1220030.
[13] Huang S, Wang B, Li X, Zheng P, Mourtzis D, Wang L. Industry 5.0 and Society 5.0—Comparison, complementation and co-evolution. J Manuf Syst 2022;64. https://doi.org/10.1016/j.jmsy.2022.07.010.
[14] Li D, Liu S, Wang B, Yu C, Zheng P, Li W. Trustworthy AI for human-centric smart manufacturing: a survey. J Manuf Syst 2025;78. https://doi.org/10.1016/j.jmsy.2024.11.020.
[15] Tian Y, Liu A. Responsible manufacturing toward Industry 5.0. Manufacturing from Industry 4.0 to Industry 5.0, 8. Elsevier Inc; 2024. p. 231–63. https://doi.org/10.1016/B978-0-443-13924-6.00008-9.
[16] R. David, J. Stahre, T. Wuest , An operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies, International Conference on Computers & Industrial Engineering CIE 46.
[17] Lorenzini M, Lagomarsino M, Fortini L, Gholami S, Ajoudani A. Ergonomic human-robot collaboration in industry: a review. Front Robot AI 2023;9. https://doi.org/10.3389/frobt.2022.813907.
[18] Li X, Koren Y, Epureanu BI. Complementary learning-team machines to enlighten and exploit human expertise. CIRP Ann 2022;71. https://doi.org/10.1016/j.cirp.2022.04.019.
[19] Wang L. A futuristic perspective on human-centric assembly. J Manuf Syst 2022;62. https://doi.org/10.1016/j.jmsy.2021.11.001.
[20] Wan PK, Leirmo TL. Human-centric zero-defect manufacturing: state-of-the-art review, perspectives, and challenges. Comput Ind 2023;144. https://doi.org/10.1016/j.compind.2022.103792.
[21] Kim DB, Bajestani MS, Lee JY, Shin S, Noh SD. Introduction of human-in-the-loop in smart manufacturing (H-SM). Int J Precis Eng Manuf-Smart Technol 2024;2. https://doi.org/10.57062/ijpem-st.2024.00115.
[22] Mayr-Dorn C, Winterer M, Salomon C, Hohensinger D, Fürschuss H. Assessing industrial end-user programming of robotic production cells: a controlled experiment. J Syst Softw 2023;195. https://doi.org/10.1016/j.jss.2022.111547.
[23] Villagrossi E, Delledonne M, Faroni M, Beschi M, Pedrocchi N. Hiding task-oriented programming complexity: an industrial case study. Int J Comput Integr Manuf 2023;36. https://doi.org/10.1080/0951192X.2023.2203676.
[24] R.H. Kambhamettu, M. Jae-Yoon Chung, V. Ranganeni, P. Alves-Oliveira, Collecting Insights about How Novice Programmers Naturally Express Programs for Robots.
[25] Nooralahzadeh F, Zhang Y, Furst J, Stockinger K. Explainable multi-modal. Data Explor Nat Lang via LLM Agent 2024. https://doi.org/10.48550/arxiv.2412.18428.
[26] Enhancing user experience in large language models through human-centered design: Integrating theoretical insights with an experimental study to meet diverse software learning needs with a single document knowledge base, arXiv.org, 2024, ⟨https://www.proquest.com/docview/3057514750⟩.
[27] Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. arXiv (Cornell University). A Compr Overv Large Lang Models 2023. https://doi.org/10.48550/arXiv.2307.06435.
[28] Leng J, Zhu X, Huang Z, Li X, Zheng P, Zhou X, et al. Unlocking the power of industrial artificial intelligence towards Industry 5.0: insights, pathways, and challenges. J Manuf Syst 2024;73. https://doi.org/10.1016/j.jmsy.2024.02.010.
[29] Heredia Álvaro JA, Barreda JG. An advanced retrieval-augmented generation system for manufacturing quality control. Adv Eng Inform 2024;64. https://doi.org/10.1016/j.aei.2024.103007.
[30] Baptista ML, Yue N, Manjurul Islam MM, Prendinger H. Large Language Models (LLMs) for Smart Manufacturing and Industry X.0. Artificial Intelligence for Smart Manufacturing and Industry X.0. Cham: Springer; 2025. p. 97–119. https://doi.org/10.1007/978-3-031-80154-9_5.
[31] Colabianchi S, Tedeschi A, Costantino F. Human-technology integration with industrial conversational agents: a conceptual architecture and a taxonomy for manufacturing. J Ind Inf Integr 2023;35. https://doi.org/10.1016/j.jii.2023.100510.
[32] Chen C, Zhao K, Leng J, Liu C, Fan J, Zheng P. Integrating large language model and digital twins in the context of industry 5.0: framework, challenges and opportunities. Robot Comput-Integr Manuf 2025:94. https://doi.org/10.1016/j.rcim.2025.102982.
[33] Wang X, Jiang Z, Xiong Y, Liu A. Human-LLM collaboration in generative design for customization. J Manuf Syst 2025;80. https://doi.org/10.1016/j.jmsy.2025.03.012.
[34] Bi J, Wang Z, Yuan H, Shi X, Wang Z, Zhang J, et al. Large AI models and their applications: classification, limitations, and potential solutions. Softw Pract & Exp 2025;55. https://doi.org/10.1002/spe.3408.
[35] Wang T, Zhang B, Jiang D, Li D. A multimodal large language model framework for intelligent perception and decision-making in smart manufacturing. Sens 2025;25. https://doi.org/10.3390/s25103072.
[36] Wang Z, Liang X, Li M, Li S, Liu J, Zheng L. Towards cognitive intelligence-enabled product design: the evolution, state-of-the-art, and future of AI-enabled product design. J Ind Inf Integr 2025;43. https://doi.org/10.1016/j.jii.2024.100759.
[37] Fang W, Chen L, Han L, Ding J. Context-aware cognitive augmented reality assembly: Past, present, and future. J Ind Inf Integr 2025;44. https://doi.org/10.1016/j.jii.2025.100780.
[38] ElMaraghy H, Monostori L, Schuh G, ElMaraghy W. Evolution and future of manufacturing systems. CIRP Ann 2021;70. https://doi.org/10.1016/j.cirp.2021.05.008.
[39] Bin Akhtar Z. Artificial intelligence (AI) within manufacturing: an investigative exploration for opportunities, challenges, future directions. Metaverse 2024;5. https://doi.org/10.54517/m.v5i2.2731.

[40] Jwo J, Lin C, Lee C. Smart technology–driven aspects for human-in-the-loop smart manufacturing. Int J Adv Manuf Technol 2021;114. https://doi.org/10.1007/s00170-021-06977-9.
[41] Anonymous. Review of machine learning technologies and artificial intelligence in modern manufacturing systems, Design and. Operation of Production Networks for Mass Personalization in the Era of Cloud Technology. United States: Elsevier; 2021. https://doi.org/10.1016/B978-0-12-823657-4.00002-6.
[42] Leng J, Sha W, Wang B, Zheng P, Zhuang C, Liu Q, et al. Industry 5.0: prospect and retrospect. J Manuf Syst 2022;65. https://doi.org/10.1016/j.jmsy.2022.09.017.
[43] Fan J, Zheng P. A vision-language-guided robotic action planning approach for ambiguity mitigation in human–robot collaborative manufacturing. J Manuf Syst 2024;74. https://doi.org/10.1016/j.jmsy.2024.05.003.
[44] Wang T, Fan J, Zheng P. An LLM-based vision and language cobot navigation approach for Human-centric smart manufacturing. J Manuf Syst 2024;75. https://doi.org/10.1016/j.jmsy.2024.04.020.
[45] Colabianchi S, Costantino F, Sabetta N. Assessment of a large language model based digital intelligent assistant in assembly manufacturing. Comput Ind 2024;162. https://doi.org/10.1016/j.compind.2024.104129.
[46] Industry 5.0 - Towards a sustainable, human-centric and resilient European industry, Publications Office of the European Union, 2021, ⟨https://data.europa.eu/doi/10.2777/308407⟩.
[47] Industry 5.0, a transformative vision for Europe – Governing systemic transformations towards a sustainable industry, 2021, ⟨https://data.europa.eu/doi/10.2777/17322⟩.
[48] ERA industrial technologies roadmap on human-centric research and innovation for the manufacturing sector, 2024, ⟨https://data.europa.eu/doi/10.2777/0266⟩.
[49] Longo F, Padovano A, Umbrello S. Value-oriented and ethical technology engineering in industry 5.0: a human-centric perspective for the design of the factory of the future. Appl Sci 2020;10.
[50] Ciampi FG, Diallo TML, Mhenni F, Choley J, Patalano S. Digital twin-based framework for an efficient execution of CPPS reconfiguration through human–robot collaboration. J Manuf Syst 2025;83. https://doi.org/10.1016/j.jmsy.2025.09.001.
[51] Wang T, Liu Z, Wang L, Li M, Wang XV. A design framework for high-fidelity human-centric digital twin of collaborative work cell in Industry 5.0. J Manuf Syst 2025;80. https://doi.org/10.1016/j.jmsy.2025.02.018.
[52] B S, R. P P, M. B S, S K. The Evolution of Large Language Model: Models, Applications and Challenges. Int Conf Curr Trends Adv Comput (ICCTAC) 2024 Int Conf Curr Trends Adv Comput (ICCTAC) 2024:1–98. https://doi.org/10.1109/ICCTAC61556.2024.10581180.
[53] Davis R, King JJ. Orig Rule-Based Syst AI Rule-Based Expert Syst MYCIN Exp Stanf Heuristic Program Proj 1984.
[54] Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, et al. A review on large language models: architectures, applications, taxonomies, open issues and challenges. IEEE Access 2024;12. https://doi.org/10.1109/access.2024.3365742.
[55] Manning CD, Schütze H. Foundations of statistical natural language processing. First ed. Cambridge, Massachusetts: The MIT Press; 1999.
[56] Reynolds DA. Gaussian mixture models. Encycl Biom 2009;741.
[57] Mikolov T, Kombrink S, Burget L, Černocký J, Khudanpur S. Extensions of recurrent neural network language model. IEEE Int Conf Acoust Speech Signal Process (ICASSP) 2011:5528–31.
[58] Wu Y. arXiv preprint. arXiv:1609.08144, Google's Neural Mach Transl Syst Bridg gap Hum Mach Transl 2016.
[59] Ashish V. Attention is all you need. Adv Neural Inf Process Syst 2017:30.
[60] G. Jawahar, B. Sagot, D. Seddah, What does BERT learn about the structure of language?, ACL 2019-57th Annual Meeting of the Association for Computational Linguistics.
[61] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. Open Blog 2019:1.
[62] Floridi L. M. Chiriatti, GPT-3: Its nature, scope, limits, and consequences. Minds Mach 2020;30.
[63] Liu Y, Zhang Y, Wang Y, Hou F, Yuan J, Tian J, et al. A survey of visual transformers. TNNLS 2024;35. https://doi.org/10.1109/TNNLS.2022.3227717.
[64] A. Radford, Improving language understanding by generative pre-training, 2018.
[65] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, , Palm: Scaling language modeling with pathways, arXiv (Cornell University), 2022. https://doi.org/10.48550/ARXIV.2204.02311.
[66] R. Thoppilan, D. De Freitas, J. Hall, N. Shazeer, A. Kulshreshtha, H. Cheng, , Lamda: language models for dialog applications, arXiv (Cornell University), 2022. https://doi.org/10.48550/arXiv.2201.08239.
[67] R. Taylor, M. Kardas, G. Cucurull, T. Scialom, A. Hartshorn, E. Saravia, , Galactica: A Large Language Model for Science, arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2211.09085.
[68] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M. Lachaux, T. Lacroix, , LLaMA: open and efficient foundation language models, arXiv (Cornell University), 2023. https://doi.org/10.48550/arXiv.2302.13971.
[69] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, , Llama 2: Open foundation and fine-tuned chat models, arXiv preprint arXiv:2307.09288, 2023.
[70] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, , Alpaca: A strong, replicable instruction-following model, Stanford Center for Research on Foundation Models.⟨https://crfm.stanford.edu/2023/03/13/alpaca.html⟩, 2023, 3.
[71] W. Chiang, Z. Li, Z. Lin, Y. Sheng, Z. Wu, H. Zhang, , Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality, See ⟨https://vicuna.lmsys.org⟩ (accessed 14 April 2023), 2023, 2.
[72] Roziere B, Gehring J, Gloeckle F, Sootla S, Gat I, Tan XE, et al. arXiv preprint. arXiv:2308.12950, Code llama Open Found Models Code 2023.
[73] Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, et al. llama 3 herd Models arXiv Prepr arXiv 2024;2407:21783.
[74] D. Guo, Q. Zhu, D. Yang, Z. Xie, K. Dong, W. Zhang, , DeepSeek-Coder: When the Large Language Model Meets Programming–The Rise of Code Intelligence, arXiv preprint arXiv:2401.14196, 2024.
[75] Han S, Wang M, Zhang J, Li D, Duan J. A review of large language models: fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges. Electronics 2024;13. https://doi.org/10.3390/electronics13245040.
[76] Snowflake Arctic: The Best LLM for Enterprise AI — Efficiently Intelligent, Truly Open. ⟨https://www.snowflake.com/en/blog/arctic-open-efficient-foundation-language-models-snowflake/⟩ (accessed Feb 4.2025).
[77] G. Team, P. Georgiev, V.I. Lei, R. Burnell, L. Bai, A. Gulati, , Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, arXiv preprint arXiv:2403.05530, 2024.
[78] K. Hayawi, S. Shahriar, A. Cross-Domain Performance Report of Open AI ChatGPT o1 Model, 2024.
[79] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, , Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, arXiv preprint arXiv:2501.12948, 2025.
[80] A. Yang, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, , Qwen2. 5 technical report, arXiv preprint arXiv:2412.15115, 2024.
[81] Mustapha KB. A survey of emerging applications of large language models for problems in mechanics, product design, and manufacturing. Adv Eng Inform 2025;64. https://doi.org/10.1016/j.aei.2024.103066.
[82] Garcia CI, DiBattista MA, Letelier TA, Halloran HD, Camelio JA. Framework for LLM applications in manufacturing. Manuf Lett 2024;41. https://doi.org/10.1016/j.mfglet.2024.09.030.
[83] Mittal S, Khan MA, Romero D, Wuest T. A critical review of smart manufacturing & Industry 4.0 maturity models: implications for small and medium-sized enterprises (SMEs). J Manuf Syst 2018;49. https://doi.org/10.1016/j.jmsy.2018.10.005.
[84] Freire SK, Panicker SS, Ruiz-Arenas S, Rusak Z, Niforatos E. A cognitive assistant for operators: AI-powered knowledge sharing on complex systems. MPRV 2023;22. https://doi.org/10.1109/MPRV.2022.3218600.
[85] Hou W, Ji Z. Comparing large language models and human programmers for generating programming code. Adv Sci 2025;12. https://doi.org/10.1002/advs.202412279.
[86] Yin Z, Sun Q, Guo Q, Wu J, Qiu X, Huang X. Do large language models know what they don. t Know 2023. https://doi.org/10.48550/arxiv.2305.18153.
[87] Ma Y, Zheng S, Yang Z, Zheng P, Leng J, Hong J. Leveraging large language models in next generation intelligent manufacturing: retrospect and prospect. J Manuf Syst 2025;82.
[88] Zhao S, Liu S, Jiang Y, Zhao B, Lv Y, Zhang J, et al. Industrial Foundation Models (IFMs) for intelligent manufacturing: a systematic review. J Manuf Syst 2025;82. https://doi.org/10.1016/j.jmsy.2025.06.011.
[89] Du K, Yang B, Xie K, Dong N, Zhang Z, Wang S, et al. LLM-MANUF: an integrated framework of fine-tuning large language models for intelligent decision-making in manufacturing. Adv Eng Inform 2025;65. https://doi.org/10.1016/j.aei.2025.103263.
[90] Leung EKH, Balan N, Lee CKH, Xie S. The generative artificial intelligence large language product design multi-model framework for manufacturing operations. J Oper Res Soc 2025.
[91] Jayanthy S, Selvaganesh M, Sabarisan KM. Generative AI solution for CNC machines and robotics code generation. Int Conf Comput Innov Eng Sustain (ICCIES 2025.
[92] Palma G, Cecchi G, Rizzo A. Large Language models for predictive maintenance in the leather tanning industry: multimodal anomaly detection in compressors. Electronics 2025;14.
[93] Wang X, Anwer N, Dai Y, Liu A. ChatGPT for design, manufacturing, and education. Procedia CIRP 2023;119.
[94] Chen H, Li S, Fan J, Duan A, Yang C, Navarro-Alarcon D, et al. Human-in-the-loop robot learning for smart manufacturing: A human-centric perspective. IEEE Trans Autom Sci Eng 2025.
[95] Korteling JE, van de Boer-Visschedijk GC, Blankendaal RA, Boonekamp RC, Eikelboom AR. Human-versus artificial intelligence. Front Artif Intell 2021;4.
[96] Zhou J, Zhou Y, Wang B, Zang J. Human–cyber–physical systems (HCPSs) in the context of new-generation intelligent manufacturing. Engineering 2019;5.
[97] Mukherjee A, Divya AB, Sivvani M, Pal SK. Cognitive intelligence in industrial robots and manufacturing. Comput Ind Eng 2024;191.
[98] Walter S. Harmonising humans and technology: Exploring the dynamics of cognitive production, artificial intelligence and social communication in cybernetic systems. Open Res Eur 2024;4.
[99] Agrawal A, Thiel R, Jain P, Singh V, Fischer M. Digital twin: where do humans fit in? Autom Constr 2023;148. https://doi.org/10.1016/j.autcon.2023.104749.
[100] Kim JY, Lester C, Yang XJ. Beyond binary decisions: evaluating the effects of AI error type on trust and performance in AI-assisted tasks. Hum Factors 2025.
[101] Bajestani MS, Mahdi MM, Mun D, Kim DB. Human and humanoid-in-the-loop (HHitL) ecosystem: an industry 5.0 perspective. Machines 2025;13. https://doi.org/10.3390/machines13060510.

[102] Chen Y, Geng Y, Zhong F, Ji J, Jiang J, Lu Z, et al. Bi-dexhands: towards human-level bimanual dexterous manipulation. IEEE Trans Pattern Anal Mach Intell 2023;46.
[103] Sheth A, Roy K, Gaur M. Neurosymbolic artificial intelligence (why, what, and how). IEEE Intell Syst 2023;38.
[104] Yang J, Wang C, Jiang B, Song H, Meng Q. Visual perception enabled industry intelligence: state of the art, challenges and prospects. IEEE Trans Ind Inform 2020;17.
[105] Wan C, Cai P, Guo X, Wang M, Matsuhisa N, Yang L, et al. An artificial sensory neuron with visual-haptic fusion. Nat Commun 2020;11.
[106] Simões AC, Pinto A, Santos J, Pinheiro S, Romero D. Designing human-robot collaboration (HRC) workspaces in industrial settings: a systematic literature review. J Manuf Syst 2022;62.
[107] Wang X, Jiang Z, Xiong Y, Liu A. Human-llm collaboration in generative design for customization. J Manuf Syst 2025;80.
[108] Gao RX, Krüger J, Merklein M, Möhring H, Váncza J. Artificial Intelligence in manufacturing: State of the art, perspectives, and future directions. CIRP Ann 2024;73.
[109] Dignum V. Ethics in artificial intelligence: introduction to the special issue. Ethics Inf Technol 2018;20.
[110] Zhang Y, Wu J, Yu F, Xu L. Moral judgments of human vs. AI agents in moral dilemmas. Behav Sci 2023;13.
[111] Telkamp JB, Anderson MH. The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. J Bus Ethics 2022;178.
[112] Braun M, Greve M, Brendel AB, Kolbe LM. Humans supervising artificial intelligence–investigation of designs to optimize error detection. J Decis Syst 2024;33.
[113] Angelopoulos A, Michailidis ET, Nomikos N, Trakadas P, Hatziefremidis A, Voliotis S, et al. Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects. Sensors 2019;20.
[114] Bajestani MS, Kim C, Lee K, Kim DB. Self-X-based secure human-cyber-physical system (SSHCPS) for autonomous manufacturing in the era of industry 5.0. Adv Eng Inform 2026;69. https://doi.org/10.1016/j.aei.2025.104054.
[115] Li Y, Zhao H, Jiang H, Pan Y, Liu Z, Wu Z, et al. Large Lang Models Manuf 2024. https://doi.org/10.48550/arxiv.2410.21418.
[116] Wu D, Zheng P, Zhao Q, Zhang S, Qi J, Hu J, et al. Empowering natural human–robot collaboration through multimodal language models and spatial intelligence: pathways and perspectives. Robot Comput Integr Manuf 2026;97. https://doi.org/10.1016/j.rcim.2025.103064.
[117] Oyekan J, Turner C, Bax M, Graf E. From ontologies to knowledge augmented large language models for automation: a decision-making guidance for achieving human–robot collaboration in Industry 5.0. Comput Ind 2025;171. https://doi.org/10.1016/j.compind.2025.104329.
[118] Zhang C, Xu Q, Yu Y, Zhou G, Zeng K, Chang F, et al. A survey on potentials, pathways and challenges of large language models in new-generation intelligent manufacturing. Robot Comput-Integr Manuf 2024:92. https://doi.org/10.1016/j.rcim.2024.102883.
[119] H. Zhang, S.S. Dereck, Z. Wang, X. Lv, K. Xu, L. Wu, , Large Scale Foundation Models for Intelligent Manufacturing Applications: A Survey, arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2312.06718.
[120] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372. https://doi.org/10.1136/bmj.n71.
[121] Rane NL, Choudhary SP, Rane J. Intelligent manufacturing through generative artificial intelligence, such as ChatGPT or bard. SSRN J 2024. https://doi.org/10.2139/ssrn.4681747.
[122] Makatura L, Foshey M, Wang B, Hähnlein F, Ma P, Deng B, et al. How can large language models help humans in design and manufacturing? part 1: elements of The LLM-enabled computational design and manufacturing pipeline. Harv Data Sci Rev 2024. https://doi.org/10.1162/99608f92.cc80fe30.
[123] Xu S, Wei Y, Zheng P, Zhang J, Yu C. LLM enabled generative collaborative design in a mixed reality environment. J Manuf Syst 2024;74.
[124] Hyunjin C. A study on the change of manufacturing design process due to the development of A.I design and 3D printing. IOP Conf Ser Mater Sci Eng 2020;727. https://doi.org/10.1088/1757-899X/727/1/012010.
[125] K. Alrashedy, P. Tambwekar, Z. Zaidi, M. Langwasser, W. Xu, M. Gombolay, Generating CAD Code with Vision-Language Models for 3D Designs, 2024. https://doi.org/10.48550/arxiv.2410.05340.
[126] Yang K, Liu H, Zhao Y, Deng T. A new design approach of hardware implementation through natural language entry. IET Collab Intell Manuf 2023;5.
[127] Yin H, Zhang Z, Liu Y. The exploration of integrating the midjourney artificial intelligence generated content tool into design systems to direct designers towards future-oriented innovation. Systems 2023;11.
[128] V. Liu, J. Vermeulen, G. Fitzmaurice, J. Matejka, 3DALL-E: Integrating text-to-image AI in 3D design workflows, Proceedings of the 2023 ACM designing interactive systems conference 1955–1977.
[129] B. Poole, A. Jain, J.T. Barron, B. Mildenhall, Dreamfusion: Text-to-3d using 2d diffusion, arXiv preprint arXiv:2209.14988, 2022.
[130] H. Jun, A. Nichol, Shap-e: Generating conditional 3d implicit functions, arXiv preprint arXiv:2305.02463, 2023.
[131] Gao J, Shen T, Wang Z, Chen W, Yin K, Li D, et al. Get3d: a generative model of high quality 3d textured shapes learned from images. Adv Neural Inf Process Syst 2022:35.
[132] Brisco R, Hay L, Dhami S. Exploring the role of text-to-image ai in concept generation. Proc Des Soc 2023;3. https://doi.org/10.1017/pds.2023.184.
[133] Z. Wang, H. Qin, Intelligent industrial production process automatic regulation system based on LLM agents, AIEA 133–137. https://doi.org/10.1109/AIEA62095.2024.10692701.
[134] Incorporating Large Language Models into Production Systems for Enhanced Task Automation and Flexibility, arXiv.org, 2024, ⟨https://www.proquest.com/docview/3079566597⟩.
[135] Z. Zhao, D. Tang, H. Zhu, Z. Zhang, K. Chen, C. Liu, , A Large Language Model-based multi-agent manufacturing system for intelligent shopfloor, arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2405.16887.
[136] J. Lim, B. Vogel-Heuser, I. Kovalenko, Large Language Model-Enabled Multi-Agent Manufacturing Systems, CASE 3940–3946. https://doi.org/10.1109/CASE59546.2024.10711432.
[137] Farooq MU, Kang G, Seo J, Bae J, Kang S, Jang YJ. DAIM-HRI: a new human-robot integration technology for industries. IEEE Int Conf Adv Robot Soc Impacts (ARSO) 2024:7. https://doi.org/10.1109/arso60199.2024.10557811.
[138] You H, Ye Y, Zhou T, Zhu Q, Du J. Robot-enabled construction assembly with automated sequence planning based on ChatGPT: RoboGPT. Buildings 2023;13. https://doi.org/10.3390/buildings13071772.
[139] H. Xiao, P. Wang, LLM A.: Human in the Loop Large Language Models Enabled A Search for Robotics, arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2312.01797.
[140] T. Yoneda, J. Fang, P. Li, H. Zhang, T. Jiang, S. Lin, , Statler: State-Maintaining Language Models for Embodied Reasoning, ICRA 15083–15091. https://doi.org/10.1109/ICRA57147.2024.10610634.
[141] Mu Y, Zhang Q, Hu M, Wang W, Ding M, Jin J, et al. Embodiedgpt: vision-language pre-training via embodied chain of thought. Adv Neural Inf Process Syst 2024:36.
[142] Huang W, Xia F, Shah D, Driess D, Zeng A, Lu Y, et al. arXiv (Cornell University). Ground Decod Guid Text Gener Ground Models Embodied Agents 2023. https://doi.org/10.48550/arxiv.2303.00855.
[143] W. Huang, F. Xia, T. Xiao, H. Chan, J. Liang, P. Florence, , Inner monologue: embodied reasoning through planning with language models, arXiv (Cornell University), 2022. https://doi.org/10.48550/arXiv.2207.05608.
[144] I. Singh, V. Blukis, A. Mousavian, A. Goyal, D. Xu, J. Tremblay, , ProgPrompt: Generating Situated Robot Task Plans using Large Language Models, ICRA 11523–11530. https://doi.org/10.1109/ICRA48891.2023.10161317.
[145] Liu X, Erkoyuncu JA, Fuh JYH, Lu WF, Li B. Knowledge extraction for additive manufacturing process via named entity recognition with LLMs. Robot Comput-Integr Manuf 2025:93. https://doi.org/10.1016/j.rcim.2024.102900.
[146] Xu Q, Zhou G, Zhang C, Chang F, Cao Y, Zhao D. Generative AI and DT integrated intelligent process planning: a conceptual framework. Int J Adv Manuf Technol 2024.
[147] Chen J, Luo H, Huang S, Zhang M, Wang G, Yan Y, et al. Autonomous human-robot collaborative assembly method driven by the fusion of large language model and digital twin. J Phys Conf Ser 2024;2832. https://doi.org/10.1088/1742-6596/2832/1/012004.
[148] Yang J, Wang Y, Wang X, Wang X, Wang X, Wang F. Generative AI empowering parallel manufacturing: building a "6s" collaborative production ecology for manufacturing 5.0. TSMC 2024;54. https://doi.org/10.1109/TSMC.2024.3349555.
[149] A Unified Industrial Large Knowledge Model Framework in Industry 4.0 and Smart Manufacturing, arXiv.org, 2024, ⟨https://www.proquest.com/docview/2905672056⟩.
[150] Y. Xia, M. Shenoy, N. Jazdi, M. Weyrich, Towards autonomous system: flexible modular production system enhanced with large language model agents, ETFA 1–8. https://doi.org/10.1109/ETFA54631.2023.10275362.
[151] A. Macaluso, N. Cote, S. Chitta, Toward Automated Programming for Robotic Assembly Using ChatGPT, 2024 IEEE International Conference on Robotics and Automation (ICRA) 17687. https://doi.org/10.1109/icra57147.2024.10610554.
[152] Gkournelos C, Konstantinou C, Makris S. An LLM-based approach for enabling seamless Human-Robot collaboration in assembly. CIRP Ann 2024;73. https://doi.org/10.1016/j.cirp.2024.04.002.
[153] Zheng P, Li C, Fan J, Wang L. A vision-language-guided and deep reinforcement learning-enabled approach for unstructured human-robot collaborative manufacturing task fulfilment. CIRP Ann 2024;73. https://doi.org/10.1016/j.cirp.2024.04.003.
[154] Stone A, Xiao T, Lu Y, Gopalakrishnan K, Lee K, Vuong Q, et al. Open-World Object Manipulation using Pre-trained Vision-Language Models. arXiv (Cornell University). 2023. https://doi.org/10.48550/arxiv.2303.00905.
[155] Chandrasekhar A, Chan J, Ogoke F, Ajenifujah O, Barati Farimani A. AMGPT: a large language model for contextual querying in additive manufacturing. Addit Manuf Lett 2024;11. https://doi.org/10.1016/j.addlet.2024.100232.
[156] Fan H, Fuh J, Lu WF, Kumar AS, Li B. Unleashing the potential of large language models for knowledge augmentation: a practical experiment on incremental sheet forming. Procedia Comput Sci 2024;232. https://doi.org/10.1016/j.procs.2024.01.125.
[157] K. Sket, D. Potocnik, M. Ficko, S. Klancnik, Enhancing G-code programming in CNC machining using ChatGPT: a comparative study of GPT-3.5 and GPT-4.0.
[158] Jignasu A, Marshall K, Ganapathysubramanian B, Balu A, Hegde C, Krishnamurthy A. Evaluating large language models for g-code debugging, manipulation, and comprehension. IEEE LLM Aided Des Workshop (LAD) 2024:1. https://doi.org/10.1109/lad62341.2024.10691700.
[159] J. Lim, S. Patel, A. Evans, J. Pimley, Y. Li, I. Kovalenko, Enhancing Human-Robot Collaborative Assembly in Manufacturing Systems Using Large Language Models, arXiv preprint arXiv:2406.01915, 2024.

[160] Boonmee A, Wongsuwan K, Sukjai P. Consult Ind Mach Faults Large Lang Models 2024.

[161] S. Suri, S.N. Das, K. Singi, K. Dey, V.S. Sharma, V. Kaulgud, Software Engineering Using Autonomous Agents: Are We There Yet?, ASE 1855–1857. https://doi.org/10.1109/ASE56229.2023.00174.

[162] Zhou B, Li X, Liu T, Xu K, Liu W, Bao J. CausalKGPT: industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing. Adv Eng Inform 2024;59.

[163] Peifeng L, Qian L, Zhao X, Tao B. Joint knowledge graph and large language model for fault diagnosis and its application in aviation assembly. IEEE Trans Ind Inform 2024.

[164] Xia Y, Xiao Z, Jazdi N, Weyrich M. Generation of asset administration shell with large language model agents: toward semantic interoperability in digital twins in the context of industry 4.0. IEEE Access 2024;12. https://doi.org/10.1109/access.2024.3415470.

[165] Gkournelos C, Konstantinou C, Angelakis P, Michalos G, Makris S. Enabling Seamless Human-Robot Collaboration in Manufacturing Using LLMs. Springer Nature Switzerland; 2024. p. 81. https://doi.org/10.1007/978-3-031-57496-2_9.

[166] Zhang Z, Chai W, Wang J. Mani-GPT: a generative model for interactive robotic manipulation. Procedia Comput Sci 2023;226. https://doi.org/10.1016/j.procs.2023.10.649.

[167] Sun S, Li C, Zhao Z, Huang H, Xu W. Leveraging large language models for comprehensive locomotion control in humanoid robots design. Biomim Intell Robot 2024;4. https://doi.org/10.1016/j.birob.2024.100187.

[168] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, , RT-1: Robotics Transformer for Real-World Control at Scale, arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2212.06817.

[169] Brohan A, Brown N, Carbajal J, Chebotar Y, Chen X, Choromanski K, et al. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. arXiv (Cornell University); 2023. https://doi.org/10.48550/arxiv.2307.15818.

[170] Ahn M, Dwibedi D, Finn C, Arenas MG, Gopalakrishnan K, Hausman K, et al. arXiv (Cornell University). Autor Embodied Found Models Large Scale Orch Robot Agents 2024. https://doi.org/10.48550/arxiv.2401.12963.

[171] Zhao W, Li L, Zhan H, Wang Y, Fu Y. Applying large language model to a control system for multi-robot task assignment. Drones 2024;8.

[172] Badini S, Regondi S, Frontoni E, Pugliese R. Assessing the capabilities of ChatGPT to improve additive manufacturing troubleshooting. Adv Ind Eng Polym Res 2023;6. https://doi.org/10.1016/j.aiepr.2023.03.003.

[173] Fan H, Liu X, Fuh JYH, Lu WF, Li B. Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics. J Intell Manuf 2024. https://doi.org/10.1007/s10845-023-02294-y.

[174] A. Jignasu, K. Marshall, B. Ganapathysubramanian, A. Balu, C. Hegde, A. Krishnamurthy, Towards Foundational AI Models for Additive Manufacturing: Language Models for G-Code Debugging, Manipulation, and Comprehension, arXiv (Cornell University), 2023. https://doi.org/10.48550/arxiv.2309.02465.

[175] A. Russell-Gilbert, A. Sommers, A. Thompson, L. Cummins, S. Mittal, S. Rahimi, , AAD-LLM: Adaptive Anomaly Detection Using Large Language Models, 2024. https://doi.org/10.48550/arxiv.2411.00914.

[176] Chen Y, Zhao Z, Liu J, Tan S, Liu C. Application of generative AI-based data augmentation technique in transformer winding deformation fault diagnosis. Eng Fail Anal 2024;159.

[177] J. Zhu, S. Cai, F. Deng, B.C. Ooi, J. Wu, Do LLMs Understand Visual Anomalies? Uncovering LLM's Capabilities in Zero-shot Anomaly Detection, Proceedings of the 32nd ACM International Conference on Multimedia 48. https://doi.org/10.1145/3664647.3681190.

[178] Xia L, Li C, Zhang C, Liu S, Zheng P. Leveraging error-assisted fine-tuning large language models for manufacturing excellence. Robot Comput-Integr Manuf 2024:88. https://doi.org/10.1016/j.rcim.2024.102728.

[179] Wu T, Li J, Bao J, Liu Q. ProcessCarbonAgent: a large language models-empowered autonomous agent for decision-making in manufacturing carbon emission management. J Manuf Syst 2024;76. https://doi.org/10.1016/j.jmsy.2024.08.008.

[180] Fan H, Liu C, Janvisloo NE, Bian S, Fuh JYH, Lu WF, et al. MaViLa: unlocking new potentials in smart manufacturing through vision language models. J Manuf Syst 2025;80. https://doi.org/10.1016/j.jmsy.2025.02.017.

[181] Li T, Wang S, Luo Y, Wan J, Luo Z, Chen M. 3-D Vision and intelligent online inspection in SMT microelectronic packaging: a review. JESTIE 2024;5. https://doi.org/10.1109/JESTIE.2024.3365030.

[182] Wang H, Li Y. Large language model empowered by domain-specific knowledge base for industrial equipment operation and maintenance. 5th Int Conf Syst Reliab Saf Eng (SRSE) 2023:474–9.

[183] Wang P, Karigiannis J, Gao RX. Ontology-integrated tuning of large language model for intelligent maintenance. CIRP Ann 2024.

[184] Naqvi SMR, Ghufran M, Meraghni S, Varnier C, Nicod J, Zerhouni N. Human knowledge centered maintenance decision support in digital twin environment. J Manuf Syst 2022;65.

[185] Naqvi SMR, Ghufran M, Varnier C, Nicod J, Javed K, Zerhouni N. Unlocking maintenance insights in industrial text through semantic search. Comput Ind 2024;157.

[186] X. Liu, J. Vatn, S. Yin, V. Maithani, Performance of ChatGPT on CMRP: Potential for Assisting Maintenance and Reliability Professionals Using Large Language Models, IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society 1–7.

[187] L. Kohl, S. Eschenbacher, P. Besinger, F. Ansari, Large language model-based chatbot for improving human-centricity in maintenance planning and operations, PHM Society European Conference 8 12.

[188] Wang H, Li C, Li Y, Tsung F. An Intelligent Industrial Visual Monitoring and Maintenance Framework Empowered by Large-scale Visual and Language Models. IEEE Trans Ind Cyber-Phys Syst 2024.

[189] Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ 2023;103. https://doi.org/10.1016/j.lindif.2023.102274.

[190] Freire S.K., Wang C., Foosherian M., Wellsandt S., Ruiz-Arenas S., Niforatos E. 2024. Knowledge Sharing in Manufacturing using Large Language Models: User Evaluation and Model Benchmarking. arXivCornell University10.48550/arxiv.2401.05200.

[191] Dimitropoulos N, Papalexis P, Michalos G, Makris S. Advancing Human-Robot Interaction Using AI – A Large Language Model (LLM) Approach. Springer Nature Switzerland; 2024. p. 116. https://doi.org/10.1007/978-3-031-57496-2_12.

[192] Fan H, Zhang H, Ma C, Wu T, Fuh JYH, Li B. Enhancing metal additive manufacturing training with the advanced vision language model: a pathway to immersive augmented reality training for non-experts. J Manuf Syst 2024;75. https://doi.org/10.1016/j.jmsy.2024.06.007.

[193] Westphal E, Seitz H. Generative artificial intelligence: analyzing its future applications in additive manufacturing. Big Data Cogn Comput 2024;8.

[194] Li Y, Starly B. Building a knowledge graph to enrich ChatGPT responses in manufacturing service discovery. J Ind Inf Integr 2024;40. https://doi.org/10.1016/j.jii.2024.100612.

[195] Wang P, Karigiannis J, Gao RX. Ontology-integrated tuning of large language model for intelligent maintenance. CIRP Ann 2024;73. https://doi.org/10.1016/j.cirp.2024.04.012.

[196] Westphal E, Seitz H. Generative artificial intelligence: analyzing its future applications in additive manufacturing. Big data Cogn Comput 2024;8. https://doi.org/10.3390/bdcc8070074.

[197] Farquhar S, Kossen J, Kuhn L, Gal Y. Detecting hallucinations in large language models using semantic entropy. Nature 2024;630. https://doi.org/10.1038/s41586-024-07421-0.

[198] Lin Z, Guan S, Zhang W, Zhang H, Li Y, Zhang H. Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. Artif Intell Rev 2024;57. https://doi.org/10.1007/s10462-024-10896-y.

[199] Sun Y, Zhang Q, Bao J, Lu Y, Liu S. Empowering digital twins with large language models for global temporal feature learning. J Manuf Syst 2024;74. https://doi.org/10.1016/j.jmsy.2024.02.015.

[200] Wu D, Qu W, Wen Y, Zhang Y, Liang B. The application, challenge, and developing trends of non-destructive testing technique for large-scale and complex engineering components fabricated by metal additive manufacturing technology in aerospace. J Nondestruct Eval 2024;43. https://doi.org/10.1007/s10921-024-01107-3.

[201] Chen C, Zhao K, Leng J, Liu C, Fan J, Zheng P. Integrating large language model and digital twins in the context of industry 5.0: Framework, challenges and opportunities. Robot Comput-Integr Manuf 2025:94. https://doi.org/10.1016/j.rcim.2025.102982.

[202] Zhang C, Xu Q, Yu Y, Zhou G, Zeng K, Chang F, et al. A survey on potentials, pathways and challenges of large language models in new-generation intelligent manufacturing. Robot Comput Integr Manuf 2025;92. https://doi.org/10.1016/j.rcim.2024.102883.

[203] Xia L, Li C, Zhang C, Liu S, Zheng P. Leveraging error-assisted fine-tuning large language models for manufacturing excellence. Robot Comput-Integr Manuf 2024:88. https://doi.org/10.1016/j.rcim.2024.102728.

[204] Dellermann D, Calma A, Lipusch N, Weber T, Weigel S, Ebel P. The future of human. AI Collab a Taxon Des Knowl Hybrid Intell Syst 2021. https://doi.org/10.48550/arxiv.2105.03354.

[205] Jarrahi MH. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Bus Horiz 2018;61. https://doi.org/10.1016/j.bushor.2018.03.007.

[206] Peruzzini M, Prati E, Pellicciari M. A framework to design smart manufacturing systems for Industry 5.0 based on the human-automation symbiosis. Int J Comput Integr Manuf 2024;37. https://doi.org/10.1080/0951192X.2023.2257634.

[207] D.B. Kim, M.S. Bajestani, G. Shao, A. Jones, S.D. Noh, Conceptual Architecture of Digital Twin With Human-in-the-Loop-Based Smart Manufacturing, Volume 3: Advanced Manufacturing. https://doi.org/10.1115/IMECE2023-112791.

[208] Park S, Humphry J. Exclusion by design: intersections of social, digital and data exclusion. Inf Commun & Soc 2019;22. https://doi.org/10.1080/1369118X.2019.1606266.

[209] Mahdi MM, Bajestani MS, Noh SD, Kim DB. Digital twin-based architecture for wire arc additive manufacturing using OPC UA. Robot Comput-Integr Manuf 2025:94. https://doi.org/10.1016/j.rcim.2024.102944.

[210] Reid T, Gibert J. Inclusion in human–machine interactions. Science 2022;375. https://doi.org/10.1126/science.abf2618.

[211] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv. Cornell University; 2020. https://doi.org/10.48550/ARXIV.2005.14165.

[212] S. Zhang, S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen, , Opt: Open pre-trained transformer language models, arXiv (Cornell University), 2022. https://doi.org/10.48550/ARXIV.2205.01068.

[213] E. Nijkamp, B. Pang, H. Hayashi, L. Tu, H. Wang, Y. Zhou, , CodeGen: An Open Large Language Model for Code with Multi-Turn Program Synthesis, arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2203.13474.

[214] Rae JW, Borgeaud S, Cai T, Millican K, Hoffmann J, Song F, et al. Scaling language models: Methods, analysis & insights from training gopher. arXiv. Cornell University; 2021. https://doi.org/10.48550/arXiv.2112.11446.

[215] Longpre S, Yauney G, Reif E, Lee K, Roberts A, Zoph B, et al. arXiv (Cornell University). A Pretrainer's Guide Train Data Meas Eff Data Age Domain Cover Qual & Toxic 2023. https://doi.org/10.48550/arxiv.2305.13169.
[216] D. Chen, Y. Huang, Z. Ma, H. Chen, X. Pan, C. Ge, , Data-Juicer: A One-Stop Data Processing System for Large Language Models, Companion of the 2024 International Conference on Management of Data 120–134. https://doi.org/10.1145/3626246.3653385.
[217] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, , Exploring the limits of transfer learning with a unified text-to-text transformer, arXiv (Cornell University), 2019. https://doi.org/10.48550/ARXIV.1910.10683.
[218] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, , BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, arXiv (Cornell University), 2019. https://doi.org/10.48550/arXiv.1910.13461.
[219] Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, et al. Scaling instruction-finetuned language models. arXiv. Cornell University; 2022. https://doi.org/10.48550/arXiv.2210.11416.
[220] B. Zhang, B. Ghorbani, A. Bapna, Y. Cheng, X. Garcia, J. Shen, , Examining Scaling and Transfer of Language Model Architectures for Machine Translation, arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2202.00528.
[221] L. Dong, Unified language model pre-training for natural language understanding and generation, 2024,. https://doi.org/10.57702/4v41nnxp, ⟨https://search.datacite.org/works/10.57702/4v41nnxp⟩.
[222] T. Wang, A. Roberts, D. Hesslow, T.L. Scao, H.W. Chung, I. Beltagy, , What Language Model Architecture and Pretraining Objective Work Best for Zero-Shot Generalization?, arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2204.05832.
[223] Y. Tay, J. Wei, H.W. Chung, V.Q. Tran, D.R. So, S. Shakeri, , Transcending Scaling Laws with 0.1% Extra Compute, arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2210.11399.
[224] A. Zeng, X. Liu, Z. Du, Z. Wang, H. Lai, M. Ding, , GLM-130B: an open bilingual pre-trained model, arXiv (Cornell University), 2022. https://doi.org/10.48550/arXiv.2210.02414.
[225] Y. Huang, Y. Cheng, A. Bapna, O. Firat, M.X. Chen, D. Chen, , Gpipe: Efficient training of giant neural networks using pipeline parallelism, arXiv (Cornell University), 2018. https://doi.org/10.48550/arXiv.1811.06965.
[226] A. Harlap, D. Narayanan, A. Phanishayee, V. Seshadri, N. Devanur, G. Ganger, , Pipedream: Fast and efficient pipeline parallel DNN training, arXiv (Cornell University), 2018. https://doi.org/10.48550/arXiv.1806.03377.
[227] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, , Mixed precision training, arXiv preprint arXiv:1710.03740, 2017.
[228] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv (Cornell University), 2014. https://doi.org/10.48550/ARXIV.1412.6980.
[229] I. Loshchilov, F. Hutter, Fixing weight decay regularization in adam, arXiv preprint arXiv:1711.05101, 2017, 5.
[230] N. Shazeer, M. Stern, Adafactor: Adaptive Learning Rates with Sublinear Memory Cost, arXiv (Cornell University), 2018. https://doi.org/10.48550/arxiv.1804.04235.
[231] M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, B. Catanzaro, Megatron-lm: Training multi-billion parameter language models using model parallelism, arXiv (Cornell University), 2019. https://doi.org/10.48550/arXiv.1909.08053.
[232] Z. Bian, Q. Xu, B. Wang, Y. You, Maximizing Parallelism in Distributed Training for Huge Neural Networks, arXiv (Cornell University), 2021. https://doi.org/10.48550/arxiv.2105.14450.
[233] B. Wang, Q. Xu, Z. Bian, Y. You, Tesseract: Parallelize the Tensor Parallelism Efficiently, Proceedings of the 51st International Conference on Parallel Processing 1–11. https://doi.org/10.1145/3545008.3545087.
[234] Q. Xu, Y. You, An Efficient 2D Method for Training Super-Large Deep Learning Models, IPDPS 222–232. https://doi.org/10.1109/IPDPS54959.2023.00031.
[235] Scao TL, Fan A, Gallé M, Webson A, Wang T, Bekman S, et al. Bloom: a 176b-parameter open-access multilingual language model. arXiv. Cornell University; 2022. https://doi.org/10.48550/arXiv.2211.05100.
[236] OpenAI, Achiam J, Adler S, Ahmad L, Aleman FL, Altenschmidt J, et al. GPT-4 technical report. arXiv. Cornell University; 2023. https://doi.org/10.48550/ARXIV.2303.08774.
[237] L. Zheng, Z. Li, H. Zhang, Y. Zhuang, Z. Chen, Y. Huang, , Alpa: Automating Inter- and Intra-Operator Parallelism for Distributed Deep Learning, arXiv (Cornell University), 2022. https://doi.org/10.48550/arxiv.2201.12023.
[238] S. Li, H. Liu, Z. Bian, J. Fang, H. Huang, Y. Liu, , Colossal-AI: A Unified Deep Learning System For Large-Scale Parallel Training, Proceedings of the 52nd International Conference on Parallel Processing 766–775. https://doi.org/10.1145/3605573.3605613.
[239] J. Rasley, S. Rajbhandari, O. Ruwase, Y. He, DeepSpeed, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 3505–3506. https://doi.org/10.1145/3394486.3406703.
[240] J. Wei, M. Bosma, V.Y. Zhao, K. Guu, A.W. Yu, B. Lester, , Finetuned language models are zero-shot learners, arXiv (Cornell University), 2021. https://doi.org/10.48550/arXiv.2109.01652.
[241] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C.L. Wainwright, P. Mishkin, , Training language models to follow instructions with human feedback, arXiv (Cornell University), 2022. https://doi.org/10.48550/ARXIV.2203.02155.
[242] V. Sanh, A. Webson, C. Raffel, S.H. Bach, L. Sutawika, Z. Alyafeai, , Multitask prompted training enables zero-shot task generalization, arXiv (Cornell University), 2021. https://doi.org/10.48550/arXiv.2110.08207.
[243] N. Muennighoff, T. Wang, L. Sutawika, A. Roberts, S. Biderman, T.L. Scao, , Crosslingual generalization through multitask finetuning, arXiv preprint arXiv:2211.01786, 2022.
[244] A. Aghajanyan, A. Gupta, A. Shrivastava, X. Chen, L. Zettlemoyer, S. Gupta, Muppet: Massive multi-task representations with pre-finetuning, arXiv preprint arXiv:2101.11038, 2021.
[245] Ibrahim N, Aboulela S, Ibrahim A, Kashef R. A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): models, evaluation metrics, benchmarks, and challenges. Discov Artif Intell 2024;4. https://doi.org/10.1007/s44163-024-00175-8.
[246] Wang L, Ma C, Feng X, Zhang Z, Yang H, Zhang J, et al. A survey on large language model based autonomous agents. Front Comput Sci 2024;18. https://doi.org/10.1007/s11704-024-40231-1.
[247] Dhanda M, Rogers BA, Hall S, Dekoninck E, Dhokia V. Reviewing human-robot collaboration in manufacturing: opportunities and challenges in the context of industry 5.0. Robot Comput-Integr Manuf 2025:93. https://doi.org/10.1016/j.rcim.2024.102937.
[248] Wu F, Shen T, Bäck T, Chen J, Huang G, Jin Y, et al. Knowledge-empowered, collaborative, and co-evolving AI models: the post-LLM roadmap. Engineering 2024. https://doi.org/10.1016/j.eng.2024.12.008.
[249] Holzinger A, Zatloukal K, Müller H. Is human oversight to AI systems still possible? New Biotechnol 2025;85. https://doi.org/10.1016/j.nbt.2024.12.003.
[250] Z. Ji, T. Yu, Y. Xu, N. Lee, E. Ishii, P. Fung, Towards Mitigating Hallucination in Large Language Models via Self-Reflection, 2023. https://doi.org/10.48550/arxiv.2310.06271.
[251] Kaur R, Klobučar T, Gabrijelčič D. Harnessing the power of language models in cybersecurity: A comprehensive review. Int J Inf Manag data Insights 2025;5. https://doi.org/10.1016/j.jjimei.2024.100315.
[252] Roy CJ, Oberkampf WL. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Comput Methods Appl Mech Eng 2011;200. https://doi.org/10.1016/j.cma.2011.03.016.
[253] Xu X, Lu Y, Vogel-Heuser B, Wang L. Industry 4.0 and Industry 5.0—Inception, conception and perception. J Manuf Syst 2021;61. https://doi.org/10.1016/j.jmsy.2021.10.006.
[254] Gautam A, Aryal MR, Deshpande S, Padalkar S, Nikolaenko M, Tang M, et al. IIoT-enabled digital twin for legacy and smart factory machines with LLM integration. J Manuf Syst 2025;80. https://doi.org/10.1016/j.jmsy.2025.03.022.
[255] Colabianchi S, Tedeschi A, Costantino F. Human-technology integration with industrial conversational agents: a conceptual architecture and a taxonomy for manufacturing. J Ind Inf Integr 2023;35. https://doi.org/10.1016/j.jii.2023.100510.
[256] S.H. Choi, K. Park, D.H. Roh, J.Y. Lee, Y. Ghasemi, H. Jeong, An XR-based Approach to Safe Human-Robot Collaboration, VRW 481–482. https://doi.org/10.1109/VRW55335.2022.00106.
[257] Madsen O, Berger U, Møller C, Heidemann Lassen A, Vejrum Waehrens B, Schou C. The Future of Smart Production for SMEs. A Methodological and Practical Approach Towards Digitalization in SMEs. 1st ed. Cham: Springer International Publishing; 2023. https://doi.org/10.1007/978-3-031-15428-7.
[258] M. Fakih, R. Dharmaji, Y. Moghaddas, G. Quiros, O. Ogundare, M.A. Al Faruque, LLM4PLC: Harnessing Large Language Models for Verifiable Programming of PLCs in Industrial Control Systems, Proceedings of the 46th International Conference on Software Engineering: Software Engineering in Practice 192. https://doi.org/10.1145/3639477.3639743.
[259] Ling S, Guo D, Rong Y, Huang GQ. Real-time data-driven synchronous reconfiguration of human-centric smart assembly cell line under graduation intelligent manufacturing system. J Manuf Syst 2022;65. https://doi.org/10.1016/j.jmsy.2022.09.022.
[260] Zhong RY, Xu X, Klotz E, Newman ST. Intelligent Manufacturing in the Context of Industry 4.0: a review. Engineering 2017;3. https://doi.org/10.1016/J.ENG.2017.05.015.
[261] K.S.D. Ka, I.D.V. Bosman, D. 2. Smit, Virtual reality in automotive digital twins: a large language model-assisted content analysis of an expert evaluation, Proceedings of the 27th International Academic Mindtrek Conference 68. https://doi.org/10.1145/3681716.3681733.
[262] A.J. Spasic, D.S. Jankovic, Using ChatGPT Standard Prompt Engineering Techniques in Lesson Preparation: Role, Instructions and Seed-Word Prompts, ICEST 47–50. https://doi.org/10.1109/ICEST58410.2023.10187269.
[263] Trivedi C, Bhattacharya P, Prasad VK, Patel V, Singh A, Tanwar S, et al. Explainable AI for Industry 5.0: vision, architecture, and potential directions. IEEE Open J Ind Appl 2024;5. https://doi.org/10.1109/ojia.2024.3399057.
[264] Deshmukh A, Sheth H, Pawar P. A Survey on Blockchain and Cryptocurrency-Based Systems. IGI Global; 2023. https://doi.org/10.4018/978-1-6684-7455-6.ch018.
[265] Gholami H. Artificial intelligence techniques for sustainable reconfigurable manufacturing systems: An AI-powered decision-making application using large language models. BDCC 2024;8. https://doi.org/10.3390/bdcc8110152.
[266] Massey R, Steele A, Marzelli M, Strauss L, Mourtil E, Schulhof K, et al. The unique and complex considerations of digital asset custody. J Secur Ope Custody 2021;13. https://doi.org/10.69554/KXEK5506.
[267] Shi D, Liedl P, Bauernhansl T. Interoperable information modelling leveraging asset administration shell and large language model for quality control toward zero defect manufacturing. J Manuf Syst 2024;77. https://doi.org/10.1016/j.jmsy.2024.10.011.
[268] Xiao Y, Zheng S, Shi J, Du X, Hong J. Knowledge graph-based manufacturing process planning: A state-of-the-art review. J Manuf Syst 2023;70. https://doi.org/10.1016/j.jmsy.2023.08.006.

[269] Sun Y, Zhang Q, Bao J, Lu Y, Liu S. Empowering digital twins with large language models for global temporal feature learning. J Manuf Syst 2024;74. https://doi.org/10.1016/j.jmsy.2024.02.015.

[270] F. Khoramnejad, E. Hossain, Generative AI for the Optimization of Next-Generation Wireless Networks: Basics, State-of-the-Art, and Open Challenges, arXiv (Cornell University), 2024. https://doi.org/10.48550/arxiv.2405.17454.

[271] Li S, Wang Z, Yan Z, Gao Y, Jiang H, Zheng P. Large language model for humanoid cognition in proactive human-robot collaboration. IEEE 20th International Conference Automation Science Engineering (CASE) 2024:540. https://doi.org/10.1109/case59546.2024.10711379.

[272] Shi D, Meyer O, Oberle M, Bauernhansl T. Dual data mapping with fine-tuned large language models and asset administration shells toward interoperable knowledge representation. Robot Comput-Integr Manuf 2024:91. https://doi.org/10.1016/j.rcim.2024.102837.

[273] Arazzi M, Nocera A, Storti E. The semioe ontology: a semantic model solution for an ioe-based industry. JIoT 2024;11. https://doi.org/10.1109/JIOT.2024.3452945.

[274] Aslanpour MS, Gill SS, Toosi AN. Performance evaluation metrics for cloud, fog and edge computing: a review, taxonomy, benchmarks and standards for future research. Internet Things 2020;12.

[275] Kang M, Lee D, Bajestani MS, Kim DB, Noh SD. Edge Computing-Based Digital Twin Framework Based on ISO 23247 for Enhancing Data Processing Capabilities. Mach (Basel) 2025;13. https://doi.org/10.3390/machines13010019.

[276] Ahn J, Yun S, Kwon J, Kim W. Literacy deep reinforcement learning-based federated digital twin scheduling for the software-defined factory. Electronics 2024;13. https://doi.org/10.3390/electronics13224452.

[277] Kim DB, Bajestani MS, Lee JY, Shin S, Kim G, Sajadieh SMM, et al. Human-in-the-loop in smart manufacturing (H-SM): a review and perspective. J Manuf Syst 2025;82. https://doi.org/10.1016/j.jmsy.2025.05.020.

[278] Dong W, Li D, Ji Y, Chen H, Liu S, Ma Z, et al. Towards a next-generation LLM empowered low-code programming industrial robotic system for human-centric smart manufacturing. J Manuf Syst 2025;83. https://doi.org/10.1016/j.jmsy.2025.10.012.

[279] Wang T, Fan J, Zheng P. An LLM-based vision and language cobot navigation approach for Human-centric Smart Manufacturing. J Manuf Syst 2024;75. https://doi.org/10.1016/j.jmsy.2024.04.020.

[280] Shafiee S. Generative AI in manufacturing: a literature review of recent applications and future prospects. Procedia CIRP 2025;132.

[281] Prasad Agrawal K. Towards adoption of generative AI in organizational settings. J Comput Inf Syst 2024;64.

[282] M. Rojahn, M. Grum, Green AI: A systematic review and meta-analysis of its definitions, lifecycle models, hardware and measurement attempts, arXiv preprint arXiv:2511.07090, 2025.

[283] Coppolino L, D'Antonio S, Mazzeo G, Uccello F. The good, the bad, and the algorithm: the impact of generative AI on cybersecurity. Neurocomputing 2025; 623. https://doi.org/10.1016/j.neucom.2025.129406.

[284] Cação J, Santos J, Antunes M. Explainable AI for industrial fault diagnosis: a systematic review. J Ind Inf Integr 2025;47. https://doi.org/10.1016/j.jii.2025.100905.

[285] Oladimeji OO, Taiwo GA, Adenusi DA. Unveiling the footprints of eXplainable AI in Industry 4.0/5.0: a systematic review and bibliometric exploration. Data Sci Manag 2025. https://doi.org/10.1016/j.dsm.2025.07.002.

[286] Besinger P, Vejnoska D, Ansari F. Responsible AI (RAI) in manufacturing: a qualitative framework. Procedia Comput Sci 2024;232. https://doi.org/10.1016/j.procs.2024.01.081.