INTERNATIONAL JOURNAL OF MODERN COMPUTER SCIENCE AND IT INNOVATIONS (IJMCSIT)

eISSN: 3087-4289

Volume. 03, Issue. 01, pp. 21-29, January 2026"

Large Language Model–Driven Digital Twins for Lean-Aware Manufacturing Execution System Optimization in Industry 4.0 Environments

Dr. Markus Vogel
Faculty of Engineering, Technical University of Munich, Germany

Article received: 01/12/2025, Article Revised: 17/12/2025, Article Accepted: 06/01/2026

ABSTRACT

The convergence of digital twins, manufacturing execution systems, and generative artificial intelligence has produced a new epistemic and technological configuration for contemporary production systems. In highly complex manufacturing environments characterized by volatile demand, high product variety, and stringent efficiency requirements, the traditional rule-based or heuristics-driven optimization of manufacturing execution systems has increasingly demonstrated its structural limitations. This article develops a comprehensive theoretical and methodological framework for understanding how large language model–driven generative artificial intelligence can be integrated with digital twin–enabled cyber-physical production systems to provide dynamic, context-aware, and lean-compatible optimization of manufacturing execution system configurations. Building on the recent conceptual and empirical insights offered by Chowdhury, Pagidoju, and Lingamgunta in their analysis of generative AI for MES optimization, this study situates LLM-based recommendation engines within the broader intellectual traditions of lean manufacturing, operations research, discrete-event simulation, and digital twin theory.

The article advances the argument that manufacturing execution systems should no longer be conceptualized merely as transactional information platforms but as adaptive cognitive infrastructures embedded in cyber-physical ecosystems. Through the integration of digital twin architectures standardized under ISO 23247 and data- and knowledge-driven modeling frameworks, MES platforms can serve as real-time operational mirrors of the shop floor. Generative AI, particularly large language models, introduces a fundamentally new layer of interpretive and configurational intelligence that allows these mirrors to not only reflect but also reason about production states, constraints, and improvement opportunities. Drawing on lean management theory, including Toyota Kata and Gemba Kaizen, the article argues that LLM-driven MES optimization can support continuous improvement by translating tacit operational knowledge into executable system configurations.

Methodologically, the article adopts a multi-layered conceptual synthesis grounded in simulation theory, cyber-physical systems design, and decision support architectures. It critically examines how NP-complete scheduling problems, traditionally addressed through algorithmic approximations and simulation-based heuristics, can be reframed through generative reasoning that dynamically explores configuration spaces. The results section develops a theoretically grounded model of how LLM-based recommendation engines, when connected to digital twins of manufacturing cells, can improve responsiveness, reduce configuration inertia, and enhance alignment between strategic objectives and shop-floor realities. The discussion situates these findings within broader debates on lean and Industry 4.0 integration, energy-aware production, supply chain coordination, and human–machine collaboration.

By synthesizing insights from manufacturing science, information systems, and artificial intelligence, this article contributes a novel conceptual architecture for intelligent MES optimization. It demonstrates that the future of manufacturing control lies not in replacing human expertise but in augmenting it through generative, explainable, and context-sensitive digital companions that operate within digital twin ecosystems.

Keywords: Digital twin manufacturing, Manufacturing execution systems, Lean production, Large language models, Cyber-physical production systems, Simulation-based optimization

INTRODUCTION

The evolution of manufacturing systems over the last century has been shaped by a persistent tension between efficiency, flexibility, and control. Early mass production systems prioritized scale and standardization, while later paradigms such as lean manufacturing emphasized waste elimination, continuous improvement, and human-centered problem solving. In the contemporary era of Industry 4.0, this tension has intensified as digitalization, connectivity, and automation have multiplied the complexity of production environments while simultaneously creating new opportunities for optimization and responsiveness. Manufacturing execution systems, which historically functioned as transactional intermediaries between enterprise planning and shop-floor control, now occupy a central position in this complexity landscape because they coordinate data, workflows, and decisions across multiple temporal and organizational layers (Kelle and Akbulut, 2005).

The growing recognition that MES platforms must evolve beyond static rule-based systems has given rise to a wide body of research on data-driven, model-based, and agent-based decision support architectures (Hilletofth and Lattila, 2012; Linnartz and Stich, 2020). At the same time, the emergence of digital twin frameworks has transformed the way manufacturing systems are conceptualized and governed. Digital twins, defined as dynamic, data-synchronized virtual representations of physical assets and processes, enable continuous monitoring, simulation, and optimization of production operations (ISO, 2020a; Park et al., 2020). By embedding MES platforms within digital twin architectures, manufacturers can create cyber-physical production systems that integrate real-time data with predictive and prescriptive analytics (Ding et al., 2019).

Despite these advances, a fundamental challenge remains unresolved: how to translate the immense volumes of heterogeneous data and the outputs of complex simulations into actionable, context-sensitive configuration decisions that align with lean principles and strategic objectives. Traditional optimization approaches, grounded in operations research and algorithmic scheduling, face inherent limitations in this regard. As Ullman (1975) and Cormen (2009) have shown, many scheduling and resource allocation problems in manufacturing belong to the class of NP-complete problems, meaning that no polynomial-time algorithm exists to solve them optimally for large instances. Consequently, MES platforms often rely on heuristics, fixed rules, or limited what-if simulations that cannot fully capture the dynamic and combinatorial nature of modern production systems.

It is within this epistemic and technological gap that generative artificial intelligence, particularly large language models, has recently emerged as a promising new paradigm for MES optimization. The study by Chowdhury, Pagidoju, and Lingamgunta (2025) represents a pivotal contribution in this area by demonstrating how LLM-driven digital manufacturing configuration recommendation systems can dynamically propose MES configurations based on contextual data, historical performance, and operational constraints. Unlike conventional optimization algorithms, large language models are capable of synthesizing structured and unstructured data, interpreting natural language inputs from human operators, and generating coherent recommendations that reflect both formal rules and tacit domain knowledge (Chowdhury et al., 2025).

The significance of this development becomes clearer when viewed through the lens of lean manufacturing theory. Lean systems, as articulated by Ruttimann (2017) and Imai (2012), are not merely collections of tools but socio-technical systems that rely on continuous learning, problem solving, and the empowerment of frontline workers. Toyota Kata, for example, emphasizes the cultivation of routines that enable organizations to adapt to changing conditions through iterative experimentation and reflection (Rother, 2009). However, as manufacturing systems become more digitally mediated, the cognitive load required to navigate and configure complex MES environments increasingly exceeds the capacity of individual human operators. This creates a paradox in which the very technologies intended to support lean operations risk undermining their human-centered ethos.

Large language models offer a potential resolution to this paradox by functioning as cognitive mediators between human intent and machine execution. By interpreting natural language descriptions of problems, goals, and constraints, LLMs can translate human reasoning into formal configuration changes within MES platforms, thereby preserving the spirit of lean problem solving while leveraging the computational power of digital systems (Chowdhury et al., 2025). When embedded within digital twin architectures, these models can also evaluate the simulated consequences of alternative configurations before they are implemented on the physical shop floor, thus reducing risk and accelerating learning cycles (Kunath and Winkler, 2018).

Despite the growing interest in generative AI for manufacturing, the academic literature remains fragmented, with limited integration between AI-driven MES optimization, digital twin theory, and lean manufacturing principles. Most existing studies focus either on technical architectures for digital twins (ISO, 2020b; Zhang et al., 2019) or on isolated applications of machine learning in production planning and logistics (Liu et al., 2020). The contribution by Chowdhury et al. (2025) is notable precisely because it bridges this gap by framing LLMs as configuration recommendation engines within digital manufacturing environments. However, their study leaves open a number of theoretical and methodological questions regarding how such systems interact with simulation models, decision support structures, and organizational routines.

The literature on simulation and modeling provides a crucial foundation for addressing these questions. Discrete-event simulation, as developed by Banks et al. (2005) and Robinson (2004), has long been used to analyze the behavior of manufacturing systems under varying conditions. Tools such as Tecnomatix Plant Simulation and AnyLogic have enabled practitioners to create detailed virtual models of production and logistics processes, which can be used to test alternative layouts, schedules, and control policies (Siderska, 2016; Borshchev, 2013). Digital twins extend this tradition by linking simulation models to real-time data streams, thereby transforming them from static analytical tools into living representations of the physical system (ISO, 2020c).

Yet even the most sophisticated simulation models do not, by themselves, solve the problem of decision making. As David and Alla (2005) and Reisig (2010) have shown in their work on Petri nets and hybrid modeling, the formal representation of system dynamics must be complemented by mechanisms for evaluating and selecting among alternative courses of action. In manufacturing contexts, these mechanisms are typically embedded in MES and decision support systems, which integrate data, models, and human judgment to guide operational choices (Kahl and Zimmer, 2017). The integration of LLMs into this architecture introduces a new form of reasoning that is neither purely algorithmic nor purely human, but rather a hybrid that can draw on vast corpora of technical knowledge and experiential data.

From a supply chain perspective, the implications of this development are equally profound. ERP and MES platforms have long been recognized as critical enablers of information sharing, coordination, and cost optimization across organizational boundaries (Kelle and Akbulut, 2005). Agent-based and data-driven decision support systems further extend this capability by allowing decentralized actors to negotiate and adapt to changing conditions (Hilletofth and Lattila, 2012; Liu et al., 2020). When LLM-driven recommendation engines are integrated into these systems, they can potentially enhance not only internal production efficiency but also external collaboration and responsiveness.

The purpose of this article is therefore to develop a comprehensive, theoretically grounded, and methodologically rigorous framework for understanding how LLM-driven digital twins can be used to optimize manufacturing execution systems in a lean-aware and Industry 4.0–compatible manner. By synthesizing insights from Chowdhury et al. (2025) with the broader literature on digital twins, simulation, lean management, and decision support, the article seeks to articulate both the opportunities and the challenges associated with this emerging paradigm. The central research question guiding this inquiry can be stated as follows: How can large language model–driven recommendation systems, when embedded in digital twin–enabled cyber-physical production systems, transform the optimization of manufacturing execution systems in ways that enhance lean performance, adaptability, and human–machine collaboration?

To answer this question, the article proceeds through a detailed methodological exposition that explicates the conceptual modeling, data integration, and reasoning processes underlying LLM-driven MES optimization. It then presents a set of analytically derived results that illustrate how such systems can influence scheduling, resource allocation, and continuous improvement processes. The discussion situates these results within ongoing scholarly debates about the relationship between lean and Industry 4.0, the role of digital twins in manufacturing governance, and the ethical and organizational implications of generative AI. Through this integrative approach, the article aims to contribute not only to the technical understanding of LLM-based MES optimization but also to the theoretical foundations of intelligent manufacturing systems.

Methodology

The methodological approach adopted in this study is grounded in a multi-layered conceptual synthesis that integrates digital twin theory, simulation modeling, manufacturing execution system architecture, and large language model–driven generative reasoning. Rather than relying on empirical experimentation in a single industrial setting, the methodology is designed to construct a theoretically coherent and analytically robust framework that can accommodate the diversity of manufacturing contexts described in the literature on cyber-physical production systems (Ding et al., 2019; Park et al., 2020). This approach is consistent with the methodological traditions of operations research and systems engineering, where conceptual models and simulation-based reasoning are often used to explore complex socio-technical systems (Banks et al., 2005; Robinson, 2004).

At the core of the methodology lies the digital twin, which serves as the epistemic bridge between physical manufacturing processes and their digital representations. According to the ISO 23247 series, a manufacturing digital twin consists of multiple layers, including the physical asset layer, the data acquisition and communication layer, the information model layer, and the application and service layer (ISO, 2020a; ISO, 2020b). In this study, the MES is positioned within the application and service layer, where it orchestrates production orders, resource assignments, and quality controls based on both real-time data and historical records. The digital twin provides the MES with a continuously updated model of the shop floor, enabling it to simulate the consequences of alternative configurations before they are enacted in the physical world (Kunath and Winkler, 2018).

To integrate large language models into this architecture, the methodology conceptualizes the LLM as a generative reasoning engine that operates at the interface between human operators, MES configuration parameters, and digital twin simulations. This conceptualization draws directly on the framework proposed by Chowdhury et al. (2025), who describe how LLMs can be used to generate configuration recommendations by interpreting operational data, performance metrics, and natural language inputs. In methodological terms, the LLM is treated not as a black-box predictor but as a probabilistic text-based model that encodes patterns of technical and managerial knowledge drawn from a wide range of manufacturing-related corpora (Chowdhury et al., 2025).

The first methodological layer involves the formal representation of manufacturing processes and MES functions within the digital twin. This representation is achieved through a combination of discrete-event simulation models, Petri nets, and data-driven process models. Discrete-event simulation provides a way to model the stochastic behavior of production systems, including machine breakdowns, setup times, and material flows (Banks et al., 2005; Robinson, 2004). Petri nets and hybrid modeling techniques add a formal structure for representing concurrent, discrete, and continuous processes, which is particularly important in complex manufacturing cells that combine automated and manual operations (David and Alla, 2005; Reisig, 2010). Data-driven models, as described by Zhang et al. (2019), capture empirical relationships between process variables, enabling the digital twin to reflect the actual performance characteristics of the physical system.

The second methodological layer concerns the integration of MES configuration parameters into the digital twin model. MES platforms typically include a wide range of configurable elements, such as dispatching rules, buffer sizes, routing policies, quality inspection frequencies, and maintenance schedules (Kelle and Akbulut, 2005). In traditional systems, these parameters are often set manually by engineers or adjusted through limited what-if analyses. In the proposed framework, however, these parameters are exposed as part of the digital twin's information model, allowing them to be dynamically modified and evaluated through simulation. This creates a configuration space that can be explored by the LLM in its role as a recommendation engine (Chowdhury et al., 2025).

The third methodological layer involves the training and deployment of the LLM within the manufacturing context. While the internal training processes of large language models are beyond the scope of this study, it is essential to specify how such models are fine-tuned and contextualized for MES optimization. Following Chowdhury et al. (2025), the methodology assumes that the LLM is augmented with domain-specific knowledge through techniques such as prompt engineering, retrieval-augmented generation, and fine-tuning on manufacturing data. This enables the model to generate recommendations that are not only linguistically coherent but also technically meaningful within the specific context of the manufacturing system.

The interaction between the LLM and the digital twin is operationalized through an iterative reasoning loop. In this loop, the LLM receives a representation of the current system state, including performance indicators, bottleneck information, and operator queries. It then generates a set of candidate MES configurations expressed in natural language and structured parameter changes. These candidates are passed to the digital twin, which simulates their impact on key performance metrics such as throughput, lead time, work-in-process, and energy consumption (Franza et al., 2017; Azevedo et al., 2012). The simulation results are fed back to the LLM, which uses them to refine its recommendations, either by selecting the most promising configuration or by generating new alternatives.

This iterative process reflects a methodological synthesis of generative AI and simulation-based optimization. Whereas traditional simulation optimization relies on algorithmic search strategies such as genetic algorithms or tabu search, the LLM-driven approach leverages probabilistic language modeling to explore the configuration space in a more semantically guided manner (Borshchev, 2013). By drawing on its embedded knowledge of manufacturing practices, lean principles, and prior case studies, the LLM can prioritize configurations that are not only numerically efficient but also operationally feasible and culturally acceptable within the organization (Rother, 2009; Imai, 2012).

A critical methodological consideration concerns the evaluation of lean performance within this framework. Lean manufacturing is traditionally assessed through a combination of quantitative metrics, such as inventory turnover and defect rates, and qualitative indicators, such as employee engagement and problem-solving culture (Ruttimann, 2017; Azevedo et al., 2012). To incorporate these dimensions into the digital twin, the methodology includes both hard data streams and soft contextual inputs. For example, operator feedback captured through MES interfaces can be fed into the LLM as natural language text, allowing it to consider human perceptions and experiences when generating recommendations (Chowdhury et al., 2025).

The methodological framework also addresses the role of supply chain and logistics considerations in MES optimization. Production systems do not operate in isolation but are embedded in networks of suppliers, warehouses, and distribution channels. Agent-based decision support models, as described by Hilletofth and Lattila (2012), provide a way to represent the interactions between these actors. In the proposed methodology, such models can be integrated into the digital twin, allowing the LLM to generate MES configurations that take into account not only internal shop-floor constraints but also external supply and demand dynamics (Liu et al., 2020).

Finally, the methodology explicitly recognizes the limitations and uncertainties inherent in both digital twins and large language models. Simulation models are always simplifications of reality, subject to modeling errors and data quality issues (Robinson, 2004). LLMs, for their part, can generate plausible but incorrect or biased recommendations if their training data or prompts are inadequate (Chowdhury et al., 2025). To mitigate these risks, the methodology incorporates human oversight and validation as integral components of the decision-making loop. Operators and engineers are expected to review, interpret, and, if necessary, override LLM-generated recommendations, thereby maintaining a balance between automation and human judgment (Rother, 2009).

Through this multi-layered methodological design, the study establishes a coherent framework for analyzing how LLM-driven digital twins can be used to optimize manufacturing execution systems in a manner that is both technologically advanced and aligned with lean principles. This framework serves as the foundation for the results and discussion that follow.

Results

The application of the proposed methodological framework yields a set of analytically derived results that illuminate the transformative potential of large language model–driven digital twins for manufacturing execution system optimization. These results are not presented as empirical measurements but as theoretically grounded interpretations of how the interaction between LLMs, digital twins, and MES architectures reshapes the dynamics of production planning, control, and continuous improvement. Each of these results is situated within the existing scholarly literature on manufacturing systems and digitalization (Chowdhury et al., 2025; Ding et al., 2019).

One of the most significant results concerns the enhanced responsiveness of MES configurations to changing production conditions. Traditional MES platforms typically rely on pre-defined rules and static parameter settings that are periodically adjusted through engineering interventions or limited what-if analyses (Kelle and Akbulut, 2005). In contrast, the LLM-driven approach enables continuous reconfiguration based on real-time data and simulated forecasts. By interpreting the evolving state of the digital twin, the LLM can propose adjustments to dispatching rules, buffer sizes, and routing policies that reflect current bottlenecks, machine availability, and demand fluctuations (Chowdhury et al., 2025; Banks et al., 2005). This dynamic adaptability represents a qualitative shift from reactive control to anticipatory optimization.

A second key result is the reduction of configuration inertia within MES platforms. Configuration inertia refers to the tendency of organizations to maintain suboptimal system settings because the cognitive and organizational costs of change are perceived as too high (Ruttimann and Stockli, 2016). The integration of LLMs lowers these costs by automating the generation and evaluation of alternative configurations. Because the digital twin can simulate the consequences of each proposed change, decision makers can see, in a transparent and evidence-based manner, how different configurations affect performance metrics such as throughput, lead time, and energy consumption (Franza et al., 2017; Chowdhury et al., 2025). This transparency fosters a more experimental and learning-oriented culture that is consistent with Toyota Kata and Gemba Kaizen principles (Rother, 2009; Imai, 2012).

The results also indicate a significant improvement in the alignment between strategic objectives and operational decisions. In many manufacturing organizations, there is a persistent gap between high-level goals, such as cost reduction or sustainability, and the day-to-day configuration of MES parameters (Azevedo et al., 2012). The LLM-driven digital twin framework bridges this gap by allowing strategic priorities to be expressed in natural language and translated into specific configuration changes. For example, a strategic directive to reduce energy consumption can be interpreted by the LLM in light of energy management models and digital twin simulations, leading to recommendations for adjusting machine schedules, maintenance windows, or process sequences (Franza et al., 2017; Chowdhury et al., 2025).

Another important result pertains to the handling of NP-complete scheduling problems. As Ullman (1975) and Cormen (2009) have shown, many of the scheduling and resource allocation challenges faced by MES platforms are computationally intractable when approached through exact optimization. The LLM-driven approach does not eliminate this complexity but reframes it by using generative reasoning to explore the configuration space in a heuristic yet semantically informed manner. By drawing on its embedded knowledge of manufacturing heuristics and best practices, the LLM can generate high-quality approximate solutions that are then validated through digital twin simulations (Borshchev, 2013; Chowdhury et al., 2025). This hybrid approach combines the strengths of human-like reasoning with the rigor of computational modeling.

The results further suggest that LLM-driven MES optimization enhances the integration of human expertise into digital manufacturing systems. Lean manufacturing theory emphasizes the importance of frontline workers' tacit knowledge and problem-solving capabilities (Imai, 2012; Rother, 2009). In the proposed framework, operators can interact with the MES through natural language interfaces, describing problems, constraints, and improvement ideas in their own words. The LLM interprets these inputs and incorporates them into its recommendation process, effectively serving as a conduit between human insight and machine execution (Chowdhury et al., 2025). This preserves the human-centered ethos of lean while leveraging the analytical power of digital twins.

From a supply chain and logistics perspective, the results indicate that LLM-driven digital twins can improve coordination and resilience. By integrating agent-based and data-driven models of supply chain interactions into the digital twin, the LLM can generate MES configurations that account for upstream and downstream constraints (Hilletofth and Lattila, 2012; Liu et al., 2020). For instance, in the face of a supplier delay, the LLM might recommend re-sequencing production orders or adjusting inventory buffers to minimize disruption, with the digital twin providing a simulated assessment of these options (Chowdhury et al., 2025).

Finally, the results reveal a new form of organizational learning enabled by the continuous feedback loop between LLMs, digital twins, and MES platforms. Each cycle of recommendation, simulation, implementation, and evaluation generates new data that can be used to refine both the digital twin model and the LLM's contextual understanding (Zhang et al., 2019; Chowdhury et al., 2025). Over time, this creates a virtuous cycle in which the manufacturing system becomes increasingly adept at anticipating and responding to change, thereby enhancing its long-term competitiveness and sustainability.

Discussion

The results presented above invite a deep theoretical and practical examination of how large language model–driven digital twins transform the nature of manufacturing execution system optimization. This transformation cannot be fully understood through a purely technical lens; rather, it must be situated within broader debates about the evolution of manufacturing paradigms, the relationship between lean and Industry 4.0, and the role of artificial intelligence in socio-technical systems (Ruttimann and Stockli, 2016; Chowdhury et al., 2025).

One of the most significant theoretical implications of this work concerns the redefinition of MES from a control-oriented information system to a cognitive infrastructure embedded in a cyber-physical ecosystem. Traditional MES architectures were designed primarily to execute predefined plans and to collect transactional data for reporting and compliance (Kelle and Akbulut, 2005). Even when augmented with advanced analytics and simulation, these systems remained largely reactive, responding to deviations after they occurred. The integration of digital twins and LLM-driven recommendation engines fundamentally alters this dynamic by enabling anticipatory, context-aware, and generative forms of control (Ding et al., 2019; Chowdhury et al., 2025).

From a lean manufacturing perspective, this shift raises both opportunities and challenges. On the one hand, the ability of LLM-driven digital twins to rapidly explore and evaluate alternative configurations aligns closely with the lean emphasis on continuous improvement and rapid experimentation (Rother, 2009; Imai, 2012). The digital twin becomes a kind of virtual gemba, where improvement ideas can be tested and refined without disrupting physical operations. On the other hand, there is a risk that excessive reliance on automated recommendations could erode the problem-solving skills and situational awareness of human operators, thereby undermining the cultural foundations of lean (Ruttimann, 2017).

The work of Chowdhury et al. (2025) offers an important counterpoint to this concern by framing LLMs not as autonomous decision makers but as collaborative partners in the configuration process. In their model, the LLM translates human intent and contextual knowledge into actionable MES configurations, while the digital twin provides a transparent and interpretable assessment of their consequences. This suggests a form of augmented lean, in which digital technologies enhance rather than replace human-centered improvement practices.

Another critical dimension of the discussion concerns the epistemological status of LLM-generated recommendations. Unlike traditional optimization algorithms, which operate on explicitly defined objective functions and constraints, LLMs generate outputs based on probabilistic patterns learned from large corpora of text and data. This raises questions about the interpretability, reliability, and accountability of their recommendations (Chowdhury et al., 2025). In manufacturing contexts, where safety, quality, and regulatory compliance are paramount, these questions cannot be ignored.

The integration of digital twins provides a partial answer by serving as a validation and explanation layer. When an LLM proposes a particular MES configuration, the digital twin can simulate its impact and provide quantitative and qualitative feedback on key performance indicators (Banks et al., 2005; Park et al., 2020). This creates a form of computational due diligence that allows human decision makers to assess the plausibility and desirability of LLM-generated recommendations. In effect, the digital twin acts as a check and balance on generative reasoning, grounding it in the modeled realities of the production system.

The discussion also intersects with broader debates about the relationship between lean and Industry 4.0. Some scholars have argued that the data-driven automation of Industry 4.0 is fundamentally at odds with the human-centered, low-tech ethos of lean (Ruttimann and Stockli, 2016). Others have suggested that the two paradigms are complementary, with digital technologies providing new tools for waste elimination and continuous improvement (Azevedo et al., 2012). The LLM-driven digital twin framework supports the latter view by demonstrating how advanced digital technologies can be harnessed to reinforce lean principles rather than supplant them.

For example, the ability of LLMs to process natural language inputs and to generate context-sensitive recommendations allows frontline workers to engage with MES platforms in a more intuitive and participatory manner (Chowdhury et al., 2025). This reduces the cognitive and technical barriers that often separate operators from digital systems, thereby fostering greater ownership and engagement. At the same time, the digital twin provides a shared, data-driven representation of the production system that can facilitate cross-functional communication and problem solving (Kunath and Winkler, 2018).

The implications for supply chain management are equally significant. In an era of globalized and volatile supply networks, the ability to rapidly reconfigure production in response to external shocks is a critical source of competitive advantage (Hilletofth and Lattila, 2012; Liu et al., 2020). LLM-driven digital twins extend this capability by integrating internal MES optimization with external logistics and supplier data. This creates a more holistic form of production planning that transcends the traditional boundaries between manufacturing, procurement, and distribution (Kelle and Akbulut, 2005).

Energy management and sustainability represent another important domain of application. As Franza et al. (2017) have shown, energy consumption is an increasingly critical factor in manufacturing performance, both from a cost and an environmental perspective. Digital twins provide a platform for modeling the energy implications of different production scenarios, while LLMs can interpret strategic sustainability goals and translate them into specific MES configurations. This creates new possibilities for aligning operational decisions with broader societal and regulatory imperatives (Chowdhury et al., 2025).

Despite these promising developments, the discussion must also acknowledge a number of limitations and unresolved challenges. One of the most significant is the quality and completeness of the data that feeds both the digital twin and the LLM. Simulation models and data-driven frameworks are only as good as the data on which they are based, and manufacturing environments are often characterized by noisy, incomplete, or biased data streams (Zhang et al., 2019; Robinson, 2004). LLMs, in turn, can propagate and even amplify these biases if they are not carefully managed and validated (Chowdhury et al., 2025).

Another challenge concerns organizational readiness and change management. The successful deployment of LLM-driven digital twins requires not only technical infrastructure but also new skills, roles, and governance structures (Linnartz and Stich, 2020). Engineers and operators must learn to interact with generative AI systems, to interpret their recommendations, and to integrate them into existing improvement routines. This process can be disruptive and may encounter resistance, particularly in organizations with deeply ingrained practices and hierarchies (Rother, 2009; Ruttimann, 2017).

Future research should therefore focus on the socio-technical dimensions of LLM-driven MES optimization, including issues of trust, accountability, and human–machine collaboration. Empirical studies in diverse industrial settings will be essential to validate and refine the theoretical framework proposed here. Moreover, there is a need for standardized methodologies and benchmarks that can be used to evaluate the performance and robustness of generative AI systems in manufacturing contexts (ISO, 2020d; Chowdhury et al., 2025).

In summary, the discussion underscores that large language model–driven digital twins represent not merely a technological innovation but a paradigm shift in the way manufacturing execution systems are conceived and governed. By integrating generative reasoning, simulation-based validation, and lean-oriented human engagement, this paradigm holds the potential to create more adaptive, resilient, and human-centered production systems.

Conclusion

This article has developed a comprehensive and theoretically grounded framework for understanding how large language model–driven digital twins can be used to optimize manufacturing execution systems in contemporary Industry 4.0 environments. By synthesizing insights from Chowdhury, Pagidoju, and Lingamgunta’s seminal work on generative AI for MES optimization with the broader literatures on digital twins, lean manufacturing, simulation, and decision support systems, the study has demonstrated that LLM-based recommendation engines have the potential to fundamentally transform the dynamics of production planning and control (Chowdhury et al., 2025).

The central contribution of the article lies in its articulation of MES platforms as cognitive infrastructures embedded within cyber-physical production systems. Through the integration of digital twin architectures and generative AI, MES platforms can move beyond static, rule-based control toward dynamic, context-aware, and anticipatory optimization. This transformation aligns closely with the principles of lean manufacturing by supporting continuous improvement, rapid experimentation, and human-centered problem solving (Rother, 2009; Imai, 2012).

At the same time, the article has highlighted the challenges and limitations associated with this emerging paradigm, including data quality issues, interpretability concerns, and organizational readiness. Addressing these challenges will require ongoing interdisciplinary research and collaboration between engineers, computer scientists, and management scholars.

Ultimately, the integration of LLM-driven digital twins into manufacturing execution systems represents a powerful new approach to navigating the complexity of modern production environments. By combining the strengths of generative reasoning, simulation-based validation, and lean-oriented human engagement, this approach offers a pathway toward more adaptive, efficient, and sustainable manufacturing systems.

References

1. Robinson, S. Simulation: The Practice of Model Development and Use; Wiley: Hoboken, NJ, USA, 2004.
2. Franza, E., Erler, F., Langer, T., Schlegel, A., Stold, J. Requirements and tasks for active energy management systems in automotive industry. Procedia Manufacturing 8 (2017) 175–182.
3. ISO. Automation Systems and Integration Digital Twin Framework for Manufacturing Part 3; International Organization for Standardization: Geneva, Switzerland, 2020.
4. Rother, M. Toyota Kata: Managing People for Improvement, Adaptiveness and Superior Results; McGraw-Hill Education: New York, NY, USA, 2009.
5. Chowdhury, P., Pagidoju, R. T., Lingamgunta, R. K. K. Generative AI for MES optimization: LLM-driven digital manufacturing configuration recommendation. International Journal of Applied Mathematics, 38(7s), 875–890, 2025.
6. Banks, J., Carson, J.S., Barry, L. Discrete-Event System Simulation, 4th ed.; Pearson: London, UK, 2005.
7. Ruttimann, B.G.; Stockli, M.T. Lean and Industry 4.0 Twins, partners, or contenders? Journal of Service Science and Management, 9, 485–500, 2016.
8. David, R., Alla, H. Discrete, Continuous, and Hybrid Petri Nets; Springer: Berlin, Germany, 2005.
9. ISO. Automation Systems and Integration Digital Twin Framework for Manufacturing Part 1; International Organization for Standardization: Geneva, Switzerland, 2020.
10. Azevedo, S.G., Govindan, K., Carvalho, H. An integrated model to assess the leanness and agility of the automotive industry. Resources, Conservation and Recycling 66 (2012) 85–94.
11. Kelle, P., Akbulut, A. The role of ERP tools in supply chain information sharing, cooperation, and cost optimization. International Journal of Production Economics 93–94 (2005) 41–52.
12. Borshchev, A. The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6; AnyLogic North America: Chicago, IL, USA, 2013.
13. Ding, K., Chan, F.T.S., Zhang, X., Zhou, G., Zhang, F. Defining a Digital Twin-based Cyber-Physical Production System. International Journal of Production Research 57 (2019) 6315–6334.
14. Ullman, J.D. NP-complete scheduling problems. Journal of Computer and System Sciences 10 (1975) 384–393.
15. ISO. Automation Systems and Integration Digital Twin Framework for Manufacturing Part 2; International Organization for Standardization: Geneva, Switzerland, 2020.
16. Zhang, C., Zhou, G., He, J., Li, Z., Cheng, W. A data- and knowledge-driven framework for digital twin manufacturing cell. Procedia CIRP 83 (2019) 345–350.
17. Ruttimann, B.G. Lean Compendium Introduction to Modern Manufacturing Theory; Springer: Berlin, Germany, 2017.
18. Cormen, T.H. Introduction to Algorithms, 3rd ed.; MIT Press: Cambridge, MA, USA, 2009.
19. Kunath, M., Winkler, H. Integrating the Digital Twin of the manufacturing system into a decision support

system. Procedia CIRP 72 (2018) 225–231.

20. Hilletofth, P., Lattila, L. Agent based decision support in the supply chain context. Industrial Management and Data Systems 112 (2012) 1217–1235.

21. ISO. Automation Systems and Integration Digital Twin Framework for Manufacturing Part 4; International Organization for Standardization: Geneva, Switzerland, 2020.

22. Siderska, J. Application of Tecnomatix Plant Simulation for Modeling Production and Logistics Processes. Business Management and Education 14 (2016) 64–73.

23. Reisig, W. Petrinetze Modellierungstechnik, Analysemethoden, Fallstudien; Vieweg: Wiesbaden, Germany, 2010.

24. Park, K.T., Lee, J., Kim, H.J., Noh, S.D. Digital twin-based cyber physical production system architectural framework. International Journal of Advanced Manufacturing Technology 106 (2020) 1787–1810.

25. Linnartz, M., Stich, V. Software-Based Assistance System for Decision Support on Supply Chain Level. In Advances in Production Management Systems; Springer: Cham, Switzerland, 2020, 209–216.

26. Liu, S., Wang, L., Wang, X.V., Wiktorsson, M. A Framework of Data-Driven Dynamic Optimisation for Smart Production Logistics. In Advances in Production Management Systems; Springer: Cham, Switzerland, 2020, 213–221.

27. Fager, P., Hanson, R., Medbo, L., Johansson, M.I. Kit preparation for mixed model assembly. Computers and Industrial Engineering 129 (2019) 169–178.

28. Aheleroff, S., Xu, X., Lu, Y., Aristizabal, M. IoT-enabled smart appliances under Industry 4.0. Computers and Electrical Engineering 87 (2020) 106772.