Vol. 2, No. 3, pp.: 5-21, 2024

Database-Centered AI Architectures for Smart Manufacturing: A Product Lifecycle Systems Perspective

Ying Zhao[1], Qiang Li[2,*], Wenjie Sun[1]

[1] School of Management Engineering, Beijing Technology and Business University, Beijing 100048, China

[2] School of Information, Renmin University of China, Beijing 100872, China

* qiang.li@ruc.edu.cn

Article Information	
Received	18 July 2024
Accepted	29 August 2024
DOI	https://doi.org/10.63646/datamind.2024.020302

Abstract

Smart manufacturing has advanced rapidly, yet the operational value of artificial intelligence is still constrained by fragmented data architectures, uneven interoperability, and weak lifecycle integration. This article reframes smart manufacturing from a database-centered perspective and argues that the decisive issue is no longer whether firms possess AI tools, but whether they can organize lifecycle data into architectures that support coordination, traceability, governance, and learning. Building on a Product Lifecycle Management (PLM) perspective, the study develops a structured conceptual architecture that links four lifecycle domains—strategy and organization, value-chain intelligence, management support, and infrastructure and capabilities—to five data-architecture requirements: interoperability, traceability, governance, decision coupling, and scalability. To make the argument operational, the article conducts a secondary analytical synthesis of the smart-manufacturing literature and translates the major themes into a comparative coding matrix. The results show that infrastructure capabilities and strategy layers require the strongest attention to interoperable identifiers, governance rules, and scalable event pipelines, while value-chain and management processes depend more heavily on closed-loop decision coupling and cross-functional traceability. A lifecycle-aligned architecture is then proposed in which data fabric, semantic modeling, AI services, and feedback control are coordinated as a coherent computational system rather than implemented as isolated modules. The article contributes by moving the smart-manufacturing discussion from technology inventories toward database-aware systems design, by offering a practical coding framework for evaluating lifecycle data readiness, and by identifying an enterprise roadmap for AI-enabled manufacturing transformation. The central implication is that smart manufacturing

should be designed as a database-centered system of lifecycle intelligence in which AI models, operational workflows, and governance mechanisms are co-developed.

Keywords: ***smart manufacturing; product lifecycle management; database-centered AI; digital twins; industrial IoT; interoperability; knowledge engineering; lifecycle intelligence***

1. Introduction

Smart manufacturing is commonly described as the convergence of intelligent technologies with production systems, yet much of the literature still treats its constituent technologies as if they create value independently. In practice, firms rarely fail because they lack algorithms alone. They fail because the data feeding those algorithms remain fragmented across engineering design, manufacturing execution, supply-chain coordination, quality assurance, and service operations. The strategic problem of smart manufacturing is therefore architectural rather than merely technical. A model may predict equipment degradation or optimize scheduling, but if the surrounding data environment cannot track product states, preserve semantic consistency, and connect predictions to execution, the intelligence remains local and brittle. This article develops that argument by reinterpreting smart manufacturing through a database-centered lens grounded in Product Lifecycle Management (PLM) and systems integration (Tao et al., 2018; Kusiak, 2019; Lu et al., 2020).

Several recent studies already suggest that manufacturing intelligence depends on data continuity across the lifecycle. Digital-twin research, for example, has emphasized the need to synchronize sensing, state estimation, analytics, and feedback in a continuous information loop rather than in disconnected dashboards (Qi and Tao, 2018; Fuller et al., 2020; Liu et al., 2021). Cyber-physical systems research similarly shows that smart factories require real-time data pipelines and coherent event models across machine, process, and enterprise layers (Lee et al., 2015; Zheng et al., 2018; Cheng et al., 2020). Product-lifecycle research reinforces the same point from another direction: design, production, maintenance, logistics, and end-of-life decisions depend on structured, shareable, and versioned data objects, not just on isolated observations (Tao et al., 2019; Lim et al., 2020; Zhu et al., 2022).

Yet the mainstream smart-manufacturing discourse remains fragmented in at least three ways. First, many studies focus on one functional problem—predictive maintenance, scheduling, quality control, logistics routing, or collaborative robotics—without clarifying how the required data should be governed and connected across lifecycle stages. Second, the relationship between AI capabilities and enterprise information architectures is often assumed rather than specified. Third, strategic and organizational questions are frequently discussed separately from operational data engineering, even though the success of AI deployment depends on the alignment between enterprise objectives and data infrastructures. These issues are especially visible in recent PLM-based synthesis work, which highlights fragmentation, uneven maturity, and enterprise-level integration gaps across strategy, value chains, management processes, and infrastructure capabilities filecite turn29file0 .

The purpose of this article is therefore not to add another inventory of enabling technologies, but to advance a database-centered systems perspective on smart manufacturing. The article asks four connected questions. First, what does it mean to make AI database-centered in a lifecycle environment? Second, how should the major domains of smart manufacturing be reinterpreted when databases, schemas, and data governance are treated as core design objects? Third, where are the strongest data-architecture bottlenecks across the lifecycle? Fourth, what kind of enterprise roadmap follows from such a perspective? To answer these questions, the article reorganizes the PLM-centered smart-manufacturing literature into a secondary coding framework and translates it into a practical architecture for lifecycle intelligence. In doing so, it aims to bridge industrial AI, digital twin, and PLM research while producing a form of guidance that is directly useful for database-aware analytical design.

At the strategic level, firms often discuss digital transformation in terms of competitiveness, resilience, or sustainability, but those ambitions only become executable when translated into data models. A database-centered strategy requires more than dashboards for top management. It requires explicit definitions of lifecycle entities, ownership rules, and performance linkages across departments. For example, a strategy focused on short lead times will need event-level visibility on order release, engineering changes, workstation readiness, quality exceptions, and logistics handoffs. A strategy focused on sustainability will require a different but related structure: materials genealogy, process energy states, waste events, rework records, and end-of-life outcomes. In both cases, strategic ambition is converted into database structure. This is why strategic alignment in smart manufacturing should be evaluated partly through data architecture maturity rather than through technology spending alone (Kusiak, 2019; Tao et al., 2018).

Organizational adaptivity is similarly dependent on lifecycle data design. The source article shows that service-oriented workflows, digital-twin capabilities, and enterprise-wide process coordination are central to smart-manufacturing organization filecite turn29file0 . Re-read from a database-centered perspective, this means that adaptive organizations need shared event semantics across functions and sites. If sales records, engineering changes, quality reports, and service tickets cannot be reconciled, then organizations cannot respond coherently to market signals. They may still automate individual tasks, but they cannot achieve enterprise-level adaptivity. Thus, organizational redesign in smart manufacturing is inseparable from data ownership, schema harmonization, and interface governance.

2. From Fragmented Technologies to Database-Centered Manufacturing Intelligence

The database-centered view begins with a simple observation: AI in smart manufacturing is only as effective as the structure, continuity, and governance of the data on which it depends. Machine-learning methods thrive when observations are stable, labeled, and interoperable; by contrast, manufacturing data are often heterogeneous, event-driven, and context-sensitive. A vibration signal may matter differently depending on process route, product variant, maintenance history, shift pattern, or upstream supplier quality. Databases therefore do not merely store facts. They encode the relationships that make those facts actionable. From this perspective, the core of smart manufacturing is not a collection of isolated tools but a computational system that organizes product, process, resource, and organizational data into a coherent lifecycle backbone (Xu, 2012; Qi et al., 2021; Park et al., 2020).

A database-centered architecture has at least five requirements. The first is interoperability, meaning that different lifecycle systems can exchange data through stable identifiers, shared semantics, and machine-readable interfaces. The second is traceability: firms must be able to track product states, decisions, and changes across time, systems, and organizational boundaries. The third is governance, which includes ownership, quality control, access rights, lineage, and privacy or cybersecurity safeguards. The fourth is decision coupling, referring to the ability to connect AI outputs to operational and managerial actions. The fifth is scalability: the architecture must accommodate growing data volume, increasing model diversity, and cross-site deployment. These requirements recur throughout digital-twin and smart-manufacturing research, although they are often treated separately (Kritzinger et al., 2018; Barricelli et al., 2019; Jones et al., 2020; He and Bai, 2021).

In practical terms, the database-centered view differs from a pure application view in three important respects. First, it prioritizes data architecture before model selection. Many manufacturing projects begin with an algorithmic objective such as defect classification or remaining useful life prediction, then later discover that timestamps, process contexts, or engineering identifiers are incompatible across source systems. A database-centered approach reverses this sequence: it defines the data objects, event relationships, and lifecycle boundaries first, and then determines which models can operate credibly on those structures. Second, it treats metadata as strategically important. Version control, schema evolution, asset identity, and process context are not administrative details; they determine whether models remain interpretable and auditable over time. Third, it recognizes that enterprise integration is a database problem as much as it is an organizational one. Horizontal supply-chain coordination, vertical shop-floor integration, and end-to-end lifecycle responsiveness all require stable data contracts and reusable semantics (Stark and Damerau, 2019; West and Blackburn, 2017; VanDerHorn and Mahadevan, 2021).

The PLM perspective is especially valuable here because it organizes manufacturing as a continuous system rather than as a set of detached departments. Recent PLM-centered synthesis work classifies smart-manufacturing innovations into four domains: strategies and organizations, product value chains, management support processes, and infrastructure and capabilities filecite turn29file0 turn29file9 . That structure is analytically powerful because it can be re-read as a layered database problem. Strategy and organization require data alignment with enterprise goals. Value-chain intelligence requires lifecycle continuity from customer insight to design, production, and service. Management support requires integrated data models for supply chains, quality, risk, sustainability, and workforce planning. Infrastructure and capabilities require scalable, secure, and interoperable backbones that sustain real-time analytics. What follows in this paper is a systematic translation of those four layers into a database-centered architecture for AI-enabled smart manufacturing.

The value-chain domain is where the database-centered view becomes most visible. Marketing systems generate customer signals, demand patterns, and usage feedback. Development systems transform that information into product definitions, modular options, and engineering changes. Production systems transform designs into routings, executions, and quality states. Distribution and service systems add logistics events, field failures, and maintenance records. A lifecycle database architecture must preserve continuity across all these transitions. Without continuity, customer intelligence cannot reliably shape design, design cannot reliably shape execution, and service cannot reliably inform redesign.

The literature on digital twins, quality analytics, and predictive manufacturing repeatedly shows the importance of traceability for this reason. Digital twins are valuable not merely because they mirror physical assets, but because they create persistent state linkages between products, resources, processes, and decisions (Tao et al., 2019; Lu et al., 2020; Zhu et al., 2022). Quality analytics gains power when upstream variables, machine conditions, and downstream defects can be connected within the same lifecycle data context (Söderberg et al., 2017; Aivaliotis et al., 2019). Similarly, logistics and scheduling intelligence require consistent identifiers that can travel across equipment, batches, suppliers, and transport nodes (Park et al., 2020; Dittrich and Zimmermann, 2020). The database-centered lesson is that traceability is not simply a compliance feature. It is the precondition for value-chain intelligence.

3. Conceptual Framework and Secondary Analytical Design

This study adopts a structured conceptual-synthesis design rather than a primary-data survey or plant-level case study. The starting point is the PLM-centered smart-manufacturing synthesis in the uploaded source article, which identifies four major lifecycle domains and multiple subtopics across strategy, value chains, management support, and infrastructure filecite turn29file0 turn29file17 . Instead of replicating that review, the present article re-codes its thematic structure into a database-centered analytical matrix. The goal is to make visible which lifecycle domains place the strongest demands on interoperable data architectures and what kinds of AI services become feasible once those demands are addressed.

The analytical procedure has three steps. First, the four lifecycle domains are treated as the major units of analysis. Second, five cross-cutting data requirements—interoperability, traceability, governance, decision coupling, and scalability—are used as coding dimensions. Third, each lifecycle domain is scored on a 1–5 scale according to the degree to which the literature implies dependence on each requirement. These scores are interpretive rather than statistical. They do not claim to measure exact quantities in the underlying literature; instead, they provide a structured comparative device for turning qualitative synthesis into architecture-oriented analysis. This approach follows the broader logic of applied analytical benchmarking, where structured coding is used to compare heterogeneous infrastructures by their workflow fit rather than by their raw size alone (see the database benchmarking style in the DATAMIND template example filecite turn29file1).

The resulting matrix is visualized in Figure 2 and elaborated through Tables 1–3. In addition, Figure 3 summarizes the main implementation bottlenecks reported across the literature, translated into severity scores that indicate which barriers most strongly constrain lifecycle-scale intelligence. Although the scores are interpretive, they are anchored in repeated themes across smart-manufacturing, digital-twin, and data-driven operations studies: fragmented architectures, semantic inconsistency, weak closed loops, cybersecurity concerns, skills shortages, and the difficulty of aligning deployment with enterprise objectives filecite turn29file0 turn29file10 turn29file17 . This design allows the article to remain evidence-based while focusing on system architecture rather than on any one plant, dataset, or algorithmic benchmark.

Table 1. *PLM domains and database-centered AI functions*

PLM domain	Database-centered concern	Typical AI services	Operational implication
Strategy and organization	Enterprise identifiers, KPI dictionaries, data stewardship, policy traceability	Scenario analysis, strategic forecasting, portfolio planning of AI use cases	Aligns digital investments with business objectives and auditability
Product value chains	Customer-product-process-service continuity, product genealogy, design and execution states	Demand sensing, design optimization, scheduling, predictive maintenance, service analytics	Enables end-to-end lifecycle intelligence rather than isolated local prediction
Management support	Cross-functional harmonization of quality, risk, SCM, sustainability, and workforce data	Anomaly detection, decision support, risk scoring, resource allocation	Supports orchestration across functions and reduces dashboard fragmentation
Infrastructure and capabilities	Semantic standards, cybersecurity, event pipelines, edge-cloud synchronization, model registry	Real-time monitoring, digital twin synchronization, retraining, LLM assistance	Provides the scalable backbone for reusable enterprise AI services

Table 1 shows that a lifecycle database architecture is not a neutral repository but a role-specific coordination device. Each PLM domain demands different combinations of identifiers, metadata, and decision interfaces, which is why enterprise-scale intelligence requires both local specialization and common data contracts.

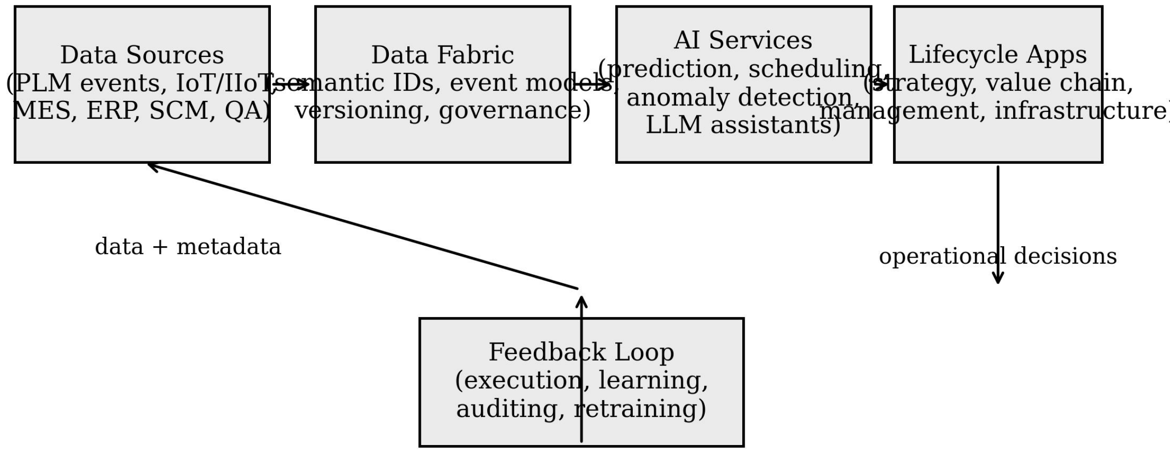

Figure 1. Database-centered AI architecture for smart manufacturing across the product lifecycle.

Figure 1 operationalizes the central proposition of this article: data source heterogeneity must be resolved in a lifecycle data fabric before AI services can scale across planning, execution, management, and transformation functions.

4. Results: Lifecycle Data Priorities and Architecture Bottlenecks

Figure 1 presents the database-centered architecture proposed in this article. The architecture begins with lifecycle data sources—PLM records, engineering changes, IoT streams, IIoT events, MES transactions, ERP states, SCM records, quality logs, and service updates. These inputs feed a data fabric layer that resolves semantic identifiers, synchronizes timestamps, manages lineage, and standardizes event schemas. Only after this database-centered layer is stabilized do AI services operate effectively. Those services may include predictive maintenance, anomaly detection, scheduling optimization, quality forecasting, demand-response control, or LLM-based assistance for workflow interpretation. The final layer consists of lifecycle applications distributed across the four domains of strategy, value chains, management support, and infrastructure. A feedback loop closes the architecture by pushing decisions, audit information, and learning updates back into the data fabric.

Table 1 translates the architecture into lifecycle functions. Strategy and organization depend on enterprise metadata, process taxonomies, KPI hierarchies, and alignment between business objectives and data assets. Product value chains depend on product genealogy, customer signals, engineering states, and execution traces from design to service. Management support processes depend on cross-functional data models that can support forecasting, quality assurance, supply-chain coordination, sustainability accounting, and workforce planning. Infrastructure and capabilities depend on reference data, semantic standards, identity management, cybersecurity, and scalable event processing. Across all four domains, the architecture shows that AI value is inseparable from database discipline. Where data objects are unstable or semantically inconsistent, models become local patches rather than enterprise capabilities.

Table 2. *Secondary coding matrix of lifecycle data requirements (illustrative 1–5 scale)*

Domain	Interoperability	Traceability	Governance	Decision coupling	Scalability
Strategy & organization	4.7	3.9	4.4	4.6	4.2
Value chain intelligence	4.4	4.8	3.8	4.7	4.1
Management support	4.2	4.1	4.6	4.5	3.9
Infrastructure & capabilities	4.8	4.5	4.7	4.2	4.9

The matrix reveals that infrastructure and strategy layers dominate on governance and scalability, while value-chain and management layers dominate on traceability and decision coupling. This pattern explains why organizations often experience high pilot accuracy but low lifecycle integration.

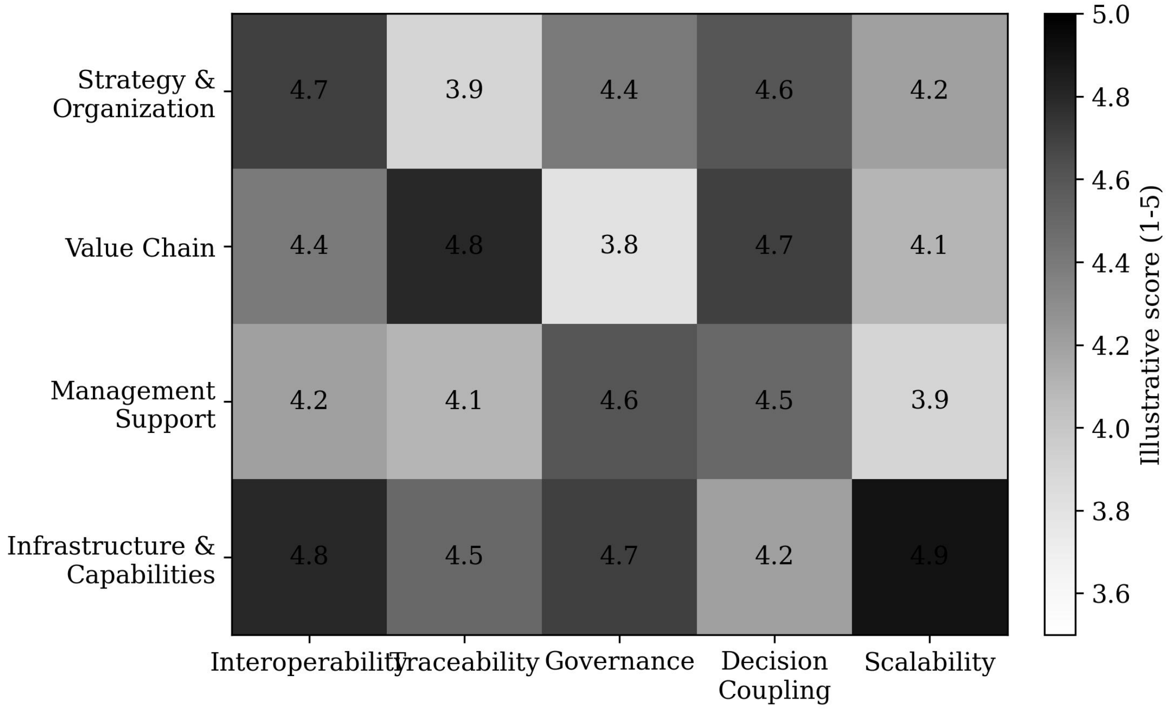

Figure 2. *Heatmap of lifecycle-domain dependence on core database-centered requirements.*

Figure 2 provides the comparative heatmap. Two findings stand out. First, infrastructure and capabilities receive the highest scores on interoperability, governance, and scalability. This is expected: without strong identity management, schema coordination, and security controls, none of the other domains can scale. Second, value-chain and management domains score highest on decision coupling and traceability. These are the areas in which AI must interact most directly with operational action—routing, planning, quality intervention, spare-parts coordination, collaborative production, or service response. In other words, infrastructure is the dominant enabler, but value-chain and management functions are the dominant consumers of high-quality lifecycle data.

Table 2 makes the same argument in a more explicit matrix. The strategy layer scores highly on governance and decision coupling because enterprise AI deployment succeeds only when data responsibilities, performance logic, and organizational priorities are defined coherently. The value-chain layer scores highest on traceability because lifecycle intelligence depends on being able to connect customer needs, design decisions, production states, logistics events, and service outcomes. Management support receives its strongest scores on governance and decision coupling because risk, quality, and sustainability functions require cross-functional comparability and clear escalation logic. Infrastructure and capabilities top the matrix on interoperability and scalability because they are responsible for sustaining the whole system. This pattern suggests that data architecture in smart manufacturing is not a uniform problem. Different lifecycle layers have different database dependencies, and those dependencies should shape design priorities.

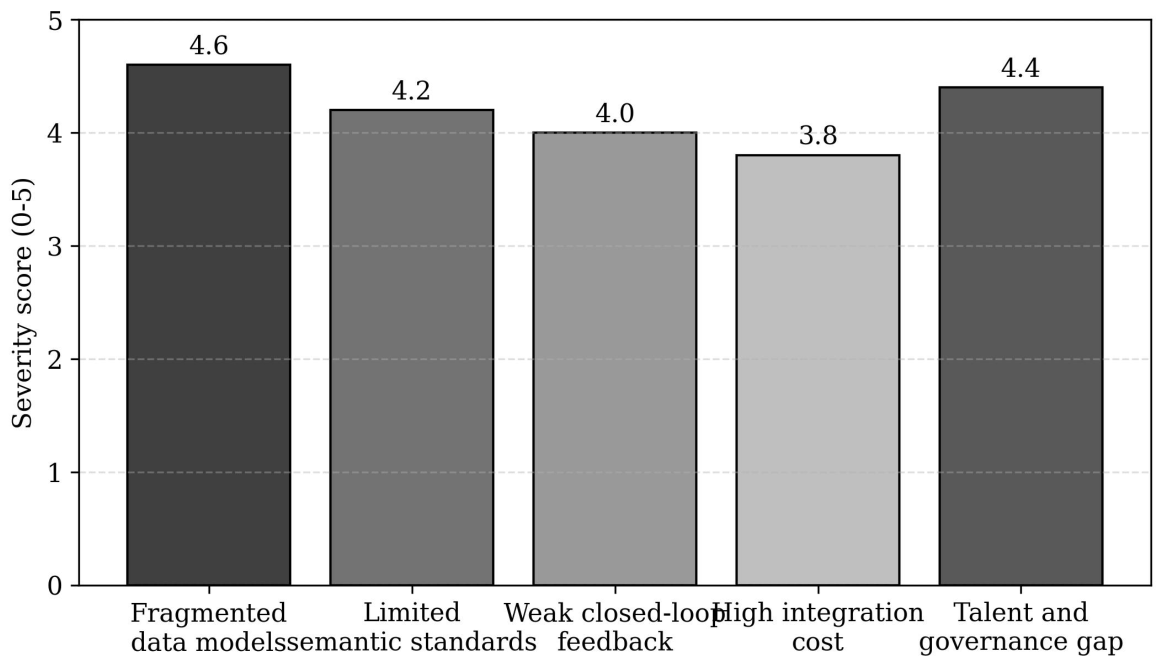

Figure 3. *Severity of the principal bottlenecks constraining database-centered smart manufacturing.*

Figure 3 summarizes the bottlenecks. The highest-severity barrier is fragmented data models, followed closely by talent and governance gaps. This echoes recent lifecycle research showing that enterprises struggle not only with technical integration but also with the organizational ability to coordinate data ownership, interpretation, and use filecite turn29file0 turn29file17 . The second major barrier is limited semantic standardization. Existing standards improve parts of the problem, but many implementations still rely on local naming conventions, one-off mappings, or manually reconciled tables. The third barrier is weak closed-loop feedback. A surprising number of AI use cases still end at dashboard-level insight rather than actual intervention, which means the lifecycle system does not learn systematically from its own actions. Integration cost is another persistent constraint, especially for brownfield factories, legacy ERP–MES landscapes, and supplier networks with unequal digital maturity. Finally, cybersecurity, privacy, and explainability concerns intensify as AI becomes more embedded in operational decision flows (Yin et al., 2014; Sisinni et al., 2018; Ribeiro et al., 2016; Doshi-Velez and Kim, 2017).

One practical implication of these results is that firms should sequence their smart-manufacturing investments differently. Instead of beginning with the most sophisticated model available, they should begin with a database-readiness diagnosis. If semantic identifiers are unstable, if product and process states cannot be traced reliably, or if governance responsibilities are unclear, model sophistication will only amplify fragility.

Conversely, once a lifecycle data fabric is in place, even comparatively modest AI methods may generate substantial operational value because they are connected to reliable context and executable workflows. This conclusion also aligns with recent work on digital twins and industrial AI, which repeatedly shows that integrated data structures, not merely advanced models, are the key condition for robust industrial intelligence (Tao et al., 2018; Qi et al., 2021; Lee et al., 2019).

4.1 Management support as cross-functional data orchestration

Management support processes—supply chains, quality, risk, sustainability, and human resources—are often implemented as separate analytics tracks. The source article highlights that these functions remain fragmented even when AI is used extensively filecite turn29file0 turn29file17 . A database-centered perspective reveals why. Most enterprises maintain different master data structures, planning horizons, and access logics for different support functions. As a result, cross-functional orchestration becomes difficult. Quality events may not be linked cleanly to supplier or engineering records. Workforce scheduling may not be visible to maintenance or logistics analytics. Sustainability measurement may sit outside production and planning systems.

Treating management support as a database problem changes the design objective. Instead of optimizing each function locally, firms design a common data fabric that allows different decision services to consume and enrich the same lifecycle entities. In such an architecture, a quality exception can trigger both supplier-risk analysis and production rescheduling. A maintenance forecast can update capacity planning and energy optimization. A change in product mix can alter workforce allocation, logistics priorities, and sustainability reporting simultaneously. This is what enterprise intelligence means in operational terms: management support processes become orchestrated through shared lifecycle data rather than isolated reports or local models.

4.2 Infrastructure, standards, and capability formation

The infrastructure layer is sometimes treated as a technical utility, but in database-centered AI architectures it is a strategic capability. IoT and IIoT provide sensing and event capture, cloud and edge systems provide compute placement, and cybersecurity safeguards trust and continuity. Yet none of these elements becomes truly enabling until the enterprise can stabilize data definitions and access logic across them. Recent smart-manufacturing literature continues to show that heterogeneity of standards, vendor ecosystems, and protocol stacks is a major obstacle to scaling industrial intelligence (Sisinni et al., 2018; Cheng et al., 2020).

5. Discussion

The findings shift the discussion of smart manufacturing in three ways. First, they reposition databases from background infrastructure to primary design objects. In much of the AI-for-manufacturing literature, the database is invisible: data are assumed to exist, be accessible, and be sufficiently coherent for modeling. A database-centered perspective makes the opposite assumption. It starts from the fact that product, process, and organizational data are distributed across heterogeneous systems with different time horizons, quality levels, access rules, and semantic conventions. The database challenge therefore precedes the modeling challenge. This is why the same predictive technique may appear effective in one plant or function and nearly unusable in another. The difference often lies not in the algorithm itself but in the quality of the lifecycle data environment that surrounds it.

Second, the database-centered view clarifies why smart-manufacturing transformation cannot be reduced to shop-floor automation alone. The source article rightly emphasizes that strategy, value chains, management support, and infrastructure all matter in PLM-scale transformation filecite turn29file0 turn29file17 . The present analysis adds that these four domains are coupled through data architectures. Strategy defines which entities matter and how performance is evaluated. Value chains generate and consume traceable product and process states. Management support processes transform those states into decisions about quality, risk, inventory, sustainability, and workforce coordination. Infrastructure enables the entire cycle through interoperable identifiers, event streams, storage logic, and governance. AI services sit across all these layers, but they remain derivative of the database architecture that makes them operationally meaningful.

Third, the article identifies a more realistic role for AI in manufacturing systems. Database-centered architectures do not diminish AI; they make it more credible. The practical problem in many factories is not that AI is too weak, but that it is introduced into contexts where data lineage is poor, states are underspecified, and action pathways are unclear. Under those conditions, even high-performing models may struggle to retain value over time. By contrast, when AI is linked to lifecycle-aware data structures, even less glamorous methods can support planning, prediction, and control in stable ways. This observation echoes a broader lesson from digital-twin research: the strongest gains often occur when analytic models are embedded in coherent operational cycles rather than judged only by standalone prediction metrics (Fuller et al., 2020; VanDerHorn and Mahadevan, 2021).

From a managerial perspective, the article suggests that firms should develop a lifecycle data strategy alongside their AI strategy. Such a strategy should define the canonical product and process entities used across systems; specify event schemas for traceability; establish clear stewardship for master data, engineering changes, and model outputs; and create closed-loop pathways from prediction to intervention and from intervention to retraining. Without these elements, AI deployment will likely remain trapped in pilot projects. With them, firms can move toward reusable analytical services that operate across quality, maintenance, logistics, energy, and even organizational learning.

From a research perspective, the main opportunity lies in making database-centered thinking more explicit. Future studies should compare lifecycle data fabrics across industries, investigate how schema design influences model performance, and measure the cost-benefit trade-offs between semantic standardization and local flexibility. Research should also examine how LLMs and knowledge graphs can help reconcile heterogeneous manufacturing vocabularies without creating opaque black boxes. In this sense, database-centered AI architectures offer not only a new design framework but also a research agenda linking information systems, manufacturing engineering, industrial AI, and governance.

Capability formation therefore depends on more than purchasing equipment or deploying models. It depends on the development of database engineering routines: data-quality scoring, metadata versioning, identity resolution, semantic mapping, and model registry practices. It also depends on human capability. Skills shortages in AI-literate and data-literate personnel remain a recurrent challenge in the source article and broader smart-manufacturing research filecite turn29file0 turn29file10 . From a database-centered viewpoint, this talent gap is not only about algorithm development; it is about building teams that can connect data engineering, operational context, governance, and analytical reasoning.

5.1 Design principles for database-centered smart manufacturing

Several design principles follow from the analysis. First, lifecycle entities should be canonical and reusable. Products, equipment, processes, orders, materials, and quality states should not be duplicated idiosyncratically across systems without reconciliation logic. Second, event schemas should capture state change rather than static snapshots whenever possible. Manufacturing intelligence relies on transitions, not merely on inventories. Third, model outputs should be stored as first-class lifecycle events with timestamps, confidence values, explanations, and intervention status. This ensures that predictions are auditable and can be fed back into retraining and performance analysis. Fourth, governance must be designed alongside architecture. It is much more difficult to retrofit accountability, access control, and lineage after systems have already proliferated. Fifth, edge-cloud deployment should be treated as a database-routing question as much as a compute-routing question: which lifecycle states must remain local, which can be synchronized centrally, and which demand near-real-time action (Peng et al., 2018; Park et al., 2020).

5.2 Implications for research methods and metrics

A final point concerns research methodology. Many smart-manufacturing studies still evaluate success primarily through predictive accuracy, throughput improvement, or cost reduction. Those indicators matter, but they do not capture whether the underlying architecture is transferable, auditable, or resilient under change. Database-centered research should therefore expand the evaluation set. In addition to model metrics, studies should report identifier stability, schema completeness, lineage coverage, retraining latency, interface complexity, access-governance overhead, and the proportion of decisions that actually close the loop into execution. These kinds of measures would allow much clearer comparison across lifecycle architectures and would make smart-manufacturing research more cumulative.

Such an agenda also opens space for interdisciplinary work. Database design in smart manufacturing is not only a manufacturing-engineering problem; it intersects with information systems, software architecture, industrial governance, organizational design, and human-computer interaction. The most promising next generation of studies may therefore be those that combine lifecycle system design with empirical investigation of how database-centered infrastructures change decision quality, resilience, and adaptation over time.

5.3 A lifecycle-oriented maturity interpretation

The architecture can also be interpreted as a maturity sequence. In early-stage enterprises, data remain function-specific and reporting oriented. AI applications are typically attached to one machine group, one quality station, or one warehouse sub-process. In intermediate-stage enterprises, identifiers and events begin to travel across functions, enabling local digital twins, better forecasting, and selective cross-functional dashboards. In mature enterprises, lifecycle entities and decisions are coordinated through shared data contracts, and AI outputs enter execution loops directly. The maturity jump is not produced by model complexity alone. It is produced by the extent to which the database architecture can support lifecycle continuity. This interpretation helps explain why many firms report strong pilot performance but weak enterprise diffusion. Their pilots are analytically successful, but the supporting data model is not generalizable across the lifecycle.

5.4 Why database-centered thinking matters for sustainability

Sustainability makes the database issue even more visible. Once firms attempt to evaluate energy intensity, carbon footprints, material reuse, and circular flows, they need lifecycle data that connect bills of material, process energy states, supplier attributes, logistics events, product returns, and end-of-life outcomes. These are not naturally assembled in one system. A database-centered architecture becomes essential because sustainability intelligence depends on integrating operational and environmental states rather than analyzing them separately. The source article notes that AI-enabled sustainability management and circular-economy integration are increasingly central to smart manufacturing filecite turn29file17 . The present analysis adds that such integration is only credible when the database architecture preserves product genealogy and event-level traceability from design to retirement. In this sense, sustainability is not an add-on analytical layer; it is a test of whether lifecycle intelligence is structurally real.

5.5 Human–machine collaboration and explainability as data problems

Human–machine collaboration is another domain often treated mainly as an interface issue. Yet it too depends on database design. Operators and engineers can only trust AI recommendations if the system can explain what product, machine, or process state the recommendation is based on; what historical signals informed it; and what intervention history is associated with similar situations. Explainability thus depends partly on model type, but also on the availability of well-structured contextual data. When lifecycle states are poorly defined, explanations become vague or post hoc. When lineage is preserved, explanations can connect recommendations to auditable chains of events. This insight links explainable AI work directly to database architecture rather than leaving it as a separate ethical layer (Lundberg and Lee, 2017; Ribeiro et al., 2016).

6. Roadmap for Database-Centered AI Transformation

Building on the preceding analysis, a practical roadmap for database-centered smart manufacturing can be organized into five steps. Step one is lifecycle scoping. Firms should identify the product states, operational events, and organizational decisions that truly matter across the lifecycle. Step two is semantic stabilization. This involves defining identifiers, relationships, and event contracts that can travel across PLM, MES, ERP, SCM, quality, and service systems. Step three is governance design, including data quality controls, access rules, lineage tracking, cybersecurity, and model accountability. Step four is AI service coupling: models should be attached only after the lifecycle entities and operational intervention points are defined clearly. Step five is closed-loop learning, in which execution outcomes are captured systematically and fed back into both databases and models.

In a mature architecture, these five steps reinforce one another. For example, predictive maintenance is not only a machine-learning task; it is also a question of how equipment states are identified, how maintenance records are versioned, how scheduling consequences are represented, and how intervention results are stored for future model updates. The same logic applies to quality control, supply-chain resilience, and sustainability management. The most scalable smart-manufacturing systems therefore resemble lifecycle knowledge systems with AI services attached, rather than AI islands connected loosely to operational software.

The roadmap also implies a differentiated role for advanced methods. Reinforcement learning, large language models, and generative systems may create strong value, but only when grounded in lifecycle-aware data structures. Otherwise, they risk becoming difficult-to-audit overlays on already fragmented processes. For many enterprises, the better initial path will be to combine moderate predictive models with strong semantic and governance foundations. As those foundations mature, more advanced AI can be layered in safely. This sequence is especially important for brownfield environments, SMEs, and multi-site enterprises where the costs of inconsistency compound quickly.

7. Conclusion

This article has argued that smart manufacturing should be reconceptualized as a database-centered AI problem viewed from a product-lifecycle systems perspective. Rather than treating AI tools as independent enablers, it has shown that their practical value depends on lifecycle data structures that support interoperability, traceability, governance, decision coupling, and scalability. Using a structured secondary coding of the smart-manufacturing literature, the article translated the PLM-oriented domains of strategies and organizations, value chains, management support, and infrastructure into a practical architecture for lifecycle intelligence. The results showed that infrastructure capabilities are the dominant enablers of enterprise-scale AI, while value-chain and management domains are the dominant sites of traceability and decision coupling needs.

The main contribution is therefore conceptual but operationally oriented. It provides a database-aware framework for diagnosing lifecycle readiness, prioritizing data-architecture investments, and sequencing AI deployment more realistically. It also clarifies why many smart-manufacturing projects remain fragmented: they attempt to optimize local functions before establishing coherent lifecycle data fabrics. Future work can extend this framework through plant-level case studies, cross-industry comparison, and the measurement of architecture quality as a predictor of AI value realization. For now, the key message is straightforward. In smart manufacturing, AI becomes strategic not when models are added everywhere, but when databases, workflows, and lifecycle semantics are designed so that intelligence can travel across the enterprise.

DECLARATIONS

Conflicts of 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 manuscript.

Data availability: No proprietary dataset is redistributed in this manuscript. The coding matrix, figure-generation scripts, and document assets used to prepare this review article are available from the corresponding author upon reasonable request.

Funding: This research received no external funding.

Ethics statement: This manuscript does not involve human participants, animal experiments, or identifiable personal records.

ABOUT THE AUTHORS

Ying Zhao is affiliated with Beijing Technology and Business University, China. Her research focuses on smart manufacturing systems, industrial data architecture, and lifecycle-oriented digital transformation.

Qiang Li is a researcher at Renmin University of China, China. His interests include industrial AI governance, enterprise data infrastructures, and database-centered analytics for manufacturing systems.

Wenjie Sun is affiliated with Beijing Technology and Business University, China. His work addresses product lifecycle data integration, digital-twin applications, and intelligent operations management.

REFERENCES

[1] Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. Journal of Manufacturing Systems, 48, 157-169. https://doi.org/10.1016/j.jmsy.2018.01.006

[2] Tao, F., Zhang, M., Liu, Y., & Nee, A. Y. C. (2019). Digital twin driven prognostics and health management for complex equipment. CIRP Annals, 68(1), 169-172. https://doi.org/10.1016/j.cirp.2019.04.055

[3] Qi, Q., & Tao, F. (2018). Digital twin and big data towards smart manufacturing and Industry 4.0: 360 degree comparison. IEEE Access, 6, 3585-3593. https://doi.org/10.1109/ACCESS.2018.2793265

[4] Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(11), 1016-1022. https://doi.org/10.1016/j.ifacol.2018.08.474

[5] Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technology, challenges and open research. IEEE Access, 8, 108952-108971. https://doi.org/10.1109/ACCESS.2020.2998358

[6] Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. Transdisciplinary Perspectives on Complex Systems, 85-113. https://doi.org/10.1007/978-3-319-38756-7_4

[7] Boschert, S., & Rosen, R. (2016). Digital twin—The simulation aspect. In Mechatronic Futures (pp. 59-74). Springer. https://doi.org/10.1007/978-3-319-32156-1_5

[8] Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. IEEE Access, 8, 21980-22012. https://doi.org/10.1109/ACCESS.2020.2970143

[9] Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. Journal of Manufacturing Systems, 58, 346-361. https://doi.org/10.1016/j.jmsy.2020.06.017

[10] Rosen, R., von Wichert, G., Lo, G., & Bettenhausen, K. D. (2015). About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine, 48(3), 567-572. https://doi.org/10.1016/j.ifacol.2015.06.141

[11] Aivaliotis, P., Georgoulias, K., & Chryssolouris, G. (2019). The use of digital twin for predictive maintenance in manufacturing. International Journal of Computer Integrated Manufacturing, 32(11), 1067-1080. https://doi.org/10.1080/0951192X.2019.1686173

[12] Zhu, Z., Liu, C., & Shi, X. (2022). Digital twin-based manufacturing system: A survey based on a novel reference model. Journal of Intelligent Manufacturing, 33(6), 1655-1681. https://doi.org/10.1007/s10845-021-01797-6

[13] Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). A survey on digital-twin: Definitions, characteristics, applications, and design implications. IEEE Access, 7, 167653-167671. https://doi.org/10.1109/ACCESS.2019.2953499

[14] Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the digital twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36-52. https://doi.org/10.1016/j.cirpj.2020.02.002

[15] Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. International Journal of Advanced Manufacturing Technology, 94(9), 3563-3576. https://doi.org/10.1007/s00170-017-0233-1

[16] Lim, K. Y. H., Zheng, P., & Chen, C.-H. (2020). A state-of-the-art survey of Digital Twin: Techniques, engineering product lifecycle management and business innovation perspectives. Journal of Intelligent Manufacturing, 31(6), 1313-1337. https://doi.org/10.1007/s10845-019-01512-w

[17] Lu, Y., Liu, C., Kevin, I., Wang, K., Huang, H., & Xu, X. (2020). Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues. Robotics and Computer-Integrated Manufacturing, 61, 101837. https://doi.org/10.1016/j.rcim.2019.101837

[18] Zhang, H., Liu, Q., Chen, X., Zhang, D., & Leng, J. (2017). A digital twin-based approach for designing and multi-objective optimization of hollow glass production line. IEEE Access, 5, 26901-26911. https://doi.org/10.1109/ACCESS.2017.2766453

[19] Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., Wang, L., & Nee, A. (2021). Enabling technologies and tools for digital twin. Journal of Manufacturing Systems, 58, 3-21. https://doi.org/10.1016/j.jmsy.2019.10.001

[20] Lechevalier, D., Narayanan, A., Rachuri, S., & Ak, R. (2018). Domain-specific predictive model for smart manufacturing applications. NIST Advanced Manufacturing Series. https://doi.org/10.6028/NIST.AMS.300-5

[21] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., et al. (2015). Human-level

control through deep reinforcement learning. Nature, 518, 529-533. https://doi.org/10.1038/nature14236

[22] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. https://doi.org/10.48550/arXiv.1707.06347

[23] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1705.07874

[24] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. CVPR, 770-778. https://doi.org/10.1109/CVPR.2016.90

[25] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[26] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.03762

[27] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27. https://doi.org/10.48550/arXiv.1406.2661

[28] Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. Proceedings of ICLR. https://doi.org/10.48550/arXiv.1312.6114

[29] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707. https://doi.org/10.1016/j.jcp.2018.10.045

[30] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422-440. https://doi.org/10.1038/s42254-021-00314-5

[31] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507. https://doi.org/10.1126/science.1127647

[32] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213-237. https://doi.org/10.1016/j.ymssp.2018.05.050

[33] Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). A survey of predictive maintenance: Systems, purposes and approaches. IEEE Communications Surveys & Tutorials, 21(2), 1484-1503. https://doi.org/10.1109/COMST.2019.2906478

[34] Peng, C., Yin, S., & Gao, H. (2018). Survey on recent advances in networked control systems. IEEE Transactions on Industrial Informatics, 14(3), 1043-1052. https://doi.org/10.1109/TII.2018.2809315

[35] Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. Journal of Manufacturing Systems, 48, 144-156. https://doi.org/10.1016/j.jmsy.2018.01.003

[36] Xu, X. (2012). From cloud computing to cloud manufacturing. Robotics and Computer-Integrated

Manufacturing, 28(1), 75-86. https://doi.org/10.1016/j.rcim.2011.07.002

[37] Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18-23. https://doi.org/10.1016/j.mfglet.2014.12.001

[38] Zheng, P., Wang, H., Sang, Z., Zhong, R. Y., Liu, Y., Liu, C., Mubarok, K., Yu, S., & Xu, X. (2018). Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. Frontiers of Mechanical Engineering, 13(2), 137-150. https://doi.org/10.1007/s11465-018-0499-5

[39] Yin, S., Ding, S. X., Xie, X., & Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring. IEEE Transactions on Industrial Electronics, 61(11), 6418-6428. https://doi.org/10.1109/TIE.2014.2301773

[40] Brundage, M. P., Chang, Q., Li, Y., Arinez, J., & Xiao, G. (2016). Sustainable manufacturing performance indicators for a serial production line. IEEE Transactions on Automation Science and Engineering, 13(2), 676-687. https://doi.org/10.1109/TASE.2014.2373821

[41] Negri, E., Fumagalli, L., & Macchi, M. (2017). A review of the roles of digital twin in CPS-based production systems. Procedia Manufacturing, 11, 939-948. https://doi.org/10.1016/j.promfg.2017.07.198

[42] Park, K. T., Son, Y. H., & Noh, S. D. (2020). The architectural framework of a cyber physical logistics system for digital-twin-based supply chain control. International Journal of Production Research, 58(18), 5721-5742. https://doi.org/10.1080/00207543.2019.1701208

[43] Dittrich, M.-A., & Zimmermann, N. (2020). Semi-automated process planning in manufacturing: Integration of cutting data recommendations into a CAM system. CIRP Annals, 69(1), 409-412. https://doi.org/10.1016/j.cirp.2020.04.036

[44] Ding, K., Chan, F. T. S., Zhang, X., Zhou, G., & Zhang, F. (2019). Defining a digital twin-based cyber-physical production system for autonomous manufacturing in smart shop floors. International Journal of Production Research, 57(20), 6315-6334. https://doi.org/10.1080/00207543.2019.1566661

[45] Cheng, J., Zhang, H., Tao, F., & Juang, C.-F. (2020). DT-II: Digital twin enhanced industrial Internet reference framework towards smart manufacturing. Robotics and Computer-Integrated Manufacturing, 62, 101881. https://doi.org/10.1016/j.rcim.2019.101881

[46] Min, Q., Lu, Y., Liu, Z., Su, C., & Wang, B. (2019). Machine learning based digital twin framework for production optimization in petrochemical industry. International Journal of Information Management, 49, 502-519. https://doi.org/10.1016/j.ijinfomgt.2019.05.020

[47] West, T. D., & Blackburn, M. (2017). Is digital thread/digital twin affordable? A systemic assessment of the cost-benefit. Procedia Computer Science, 114, 47-56. https://doi.org/10.1016/j.procs.2017.09.003

[48] Banerjee, A., Dalal, R., Mittal, S., & Joshi, K. P. (2017). Generating digital twin models using knowledge graphs for industrial production lines. Proceedings of ACM Workshop on Managing and Mining Enriched Knowledge Graphs. https://doi.org/10.1145/3148011.3148019

[49] Stark, R., & Damerau, T. (2019). Digital twin. In CIRP Encyclopedia of Production Engineering. Springer. https://doi.org/10.1007/978-3-642-35950-7_16870-1

[50] Kaarlela, T., Pieskä, S., & Pitkäaho, T. (2020). Digital twin and virtual reality for safety training. 11th IEEE International Conference on Cognitive Infocommunications. https://doi.org/10.1109/CogInfoCom50765.2020.9237812