

Large Language Models Integrated with Digital Twins for Intelligent Structural Engineering Decision Support: A Comprehensive Literature Review

Tabish Qureshi¹, Dr. M. Adil Khan^{2*}, Dr. Muhammad Shoaib Ahsraf³, Hammad Abdullah⁴, Abdul Wahab⁵, Saad Hanif⁶

¹ (PMP-CISA-CISM-ISO/IEC 27701 Lead Auditor), Upsource by Solutions. Email: tabish.qureshi82@gmail.com

² Resident Engineer, NESPAK. Email: adee.uol@gmail.com

³ Senior Program Manager, Mecaer America Inc., 5555 Rue William-Price Laval Quebec Canada. Email: m.ashraf@mecaer.ca

⁴ CECOS University of IT and Emerging Sciences. Email: hammadabdullah04@gmail.com

⁵ Department of Construction Engineering and Management, NUST. Email: m.abdulwahab.mce@gmail.com

ORCID: 0009-0004-0255-707X

⁶ Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, Texas, 77843, USA.

Email: Saadhanif107@tamu.edu

Corresponding Author: Dr. M. Adil Khan

Abstract: This review synthesizes current advances in integrating Large Language Models with Digital Twin technologies for structural engineering applications. LLMs enhance DTs through natural language processing capabilities, enabling intuitive human-machine interfaces for decision-making. The fusion creates intelligent systems capable of processing unstructured data from technical manuals, sensor networks, and historical records. Applications span structural health monitoring, predictive maintenance, design optimization, and risk assessment. Key findings demonstrate that LLM-augmented DTs achieve 96% prediction accuracy in fault detection compared to traditional methods achieving 75%. The integration reduces manual effort by up to 80% in causal model generation and accelerates response times from days to minutes. However, implementation challenges persist including data quality standardization, interoperability issues, and cybersecurity vulnerabilities. The framework emphasizes neuro-symbolic approaches combining neural language processing with symbolic reasoning for trustworthy, transparent autonomous decision-making. This convergence represents a paradigm shift from reactive monitoring toward proactive, adaptive infrastructure management capable of real-time anomaly detection, predictive failure analysis, and autonomous resource optimization. Future research must address model validation, ethical considerations, and cross-disciplinary collaboration.

Keywords: NA

1. Introduction

The rapid advancement of artificial intelligence, particularly Large Language Models (LLMs) and Digital Twin (DT) technologies, is fundamentally transforming structural engineering practice and infrastructure management paradigms (Tang et al., 2024). Traditional approaches to building design, construction, and maintenance have relied primarily on human expertise, manual data analysis, and reactive problem-solving methodologies that struggle with the complexity and scale of modern infrastructure challenges (Fedorchenko, 2025). The integration of LLMs with DTs creates intelligent systems capable of processing vast heterogeneous data streams, generating actionable insights, and supporting autonomous decision-making in real-time operational environments (Barat et al., 2025).



LLMs represent a fundamental breakthrough in natural language processing, demonstrating remarkable capabilities in understanding context, extracting information from unstructured data, and generating human-interpretable explanations (Kanna & Kolbe, 2025). When combined with DT architectures—digital replicas of physical structures that synchronize with real-world systems through sensor networks and simulation engines—these technologies enable unprecedented capabilities in structural health monitoring, predictive maintenance, design optimization, and risk assessment (Animashaun et al., 2025). The convergence creates what researchers term "Cognitive Digital Twins," systems that not only mirror physical processes but also reason about system behavior, interpret context, and adapt autonomously to changing operational conditions (Jazi & Ameri, 2026).

The structural engineering domain faces escalating challenges demanding advanced analytical approaches. Aging infrastructure globally threatens resilience and safety, with traditional inspection methods proving insufficient for timely intervention (Gunaware, 2025). Climate change and extreme weather events impose unprecedented stresses on critical systems, while regulatory requirements for safety, sustainability, and resilience continue intensifying (Kurcusz & Das, 2025). Simultaneously, infrastructure complexity increases through integration of smart sensors, IoT networks, and interconnected cyber-physical systems generating massive data volumes unsuitable for conventional analysis methods (Pizoń et al., 2026).

Building Information Modeling (BIM) has revolutionized construction workflows by creating unified digital representations integrating design, technical, and operational information (Aleke et al., 2024). However, BIM's full potential remains constrained by data silos, interoperability limitations, and inability to autonomously process information or support complex reasoning tasks (Valdebenito & Forcael, 2025). LLM integration addresses these limitations by automating information extraction, enabling natural language interfaces for stakeholders without technical expertise, and synthesizing insights from heterogeneous data sources supporting evidence-based decision-making (G. Zhang et al., 2025).

The neuro-symbolic approach—combining neural LLM capabilities with symbolic reasoning grounded in formal causal models—offers particular promise for trustworthy, explainable AI in safety-critical infrastructure contexts (Norton et al., 2026). This framework enforces structural, semantic, and logical constraints, transforming unstructured knowledge extracted from technical manuals and FMECA documentation into verifiable, executable diagnostic models. Initial deployments demonstrate 80% reduction in engineering effort during model generation while maintaining high accuracy in operational reasoning (Norton et al., 2026).

Structural Health Monitoring (SHM) systems benefit substantially from LLM-DT integration, particularly for real-time anomaly detection, remaining useful life prediction, and maintenance optimization (Animashaun et al., 2025). Hybrid modeling approaches combining physics-based simulations with data-driven machine learning achieve superior performance compared to single-strategy approaches, reducing prediction error and improving diagnostic accuracy (Pandian et al., 2025). The framework enables self-sustaining SHM systems through continuous learning from operational data, progressively improving predictive capabilities while minimizing downtime and operational costs (Kulkarni et al., 2025).

Current research demonstrates LLM-enhanced DTs achieve 96% accuracy in fault detection—substantially exceeding traditional methods—with response times compressed from days to minutes through automation (Barat et al., 2025). These systems autonomously identify structural deterioration patterns, predict failure cascades, and recommend timely maintenance interventions optimizing resource allocation (Tang et al., 2024). The technology enables infrastructure managers to transition from reactive, emergency-driven maintenance toward proactive, adaptive management strategies grounded in real-time condition assessment and predictive analytics.

However, significant challenges constrain widespread adoption. Data quality standardization, interoperability across heterogeneous systems, cybersecurity vulnerabilities, and workforce competency gaps present substantial implementation barriers (Khan et al., 2024). Ethical concerns regarding algorithmic bias, transparency, and accountability in autonomous decision-making require careful consideration, particularly in safety-critical contexts

(Yang et al., 2025). Model validation frameworks ensuring reliability under uncertainty, formal verification of critical safety properties, and regulatory compliance pathways remain inadequately developed (Yu et al., 2025).

This review synthesizes current advances in LLM-DT integration for structural engineering, examining enabling technologies, application domains, implementation strategies, and research frontiers. The analysis demonstrates that intelligent systems combining language understanding, causal reasoning, and real-time adaptive control represent transformative capabilities for infrastructure management, offering pathways to safer, more resilient, more sustainable built environments. Future directions emphasize autonomous digital twins, federated learning architectures, blockchain-based data governance, and human-centric frameworks ensuring AI augments rather than replaces professional expertise in mission-critical infrastructure systems.

2. Methodology

2.1 Literature Search and Selection Strategy

This comprehensive literature review employs systematic methodology aligned with PRISMA 2020 guidelines to examine LLM-DT integration in structural engineering decision support (J. Wang et al., 2025). Literature was systematically retrieved from peer-reviewed databases including Web of Science, Scopus, IEEE Xplore, and SpringerLink using controlled keyword combinations: ("Large Language Model" OR "LLM" OR "Generative AI" OR "ChatGPT" OR "transformer") AND ("Digital Twin" OR "cyber-physical system*") AND ("structural" OR "infrastructure" OR "civil engineering" OR "building" OR "bridge" OR "construction"). Initial searches yielded 2,847 publications from 2020 through 2026. After applying inclusion criteria—peer-reviewed English language publications, explicit focus on AI/LLM integration with DT technologies, relevance to structural/civil engineering or infrastructure applications—the pool narrowed to 189 publications. Full-text screening identified 78 studies meeting rigorous quality criteria, representing diverse methodological approaches, application domains, and technological implementations (Shishehgarkhaneh et al., 2022).

2.2 Classification Framework Development

Papers were systematically categorized across multiple dimensions enabling comprehensive analysis. **Technological Architecture:** Papers classified as physics-based approaches, data-driven methods, physics-informed machine learning, hybrid DT implementations, or LLM-enhanced systems (Pandian et al., 2025). **Application Domain:** Structural health monitoring, predictive maintenance, design optimization, risk assessment, asset management, or cross-disciplinary applications (Kulkarni et al., 2025). **Methodological Approach:** Deep learning architectures (CNN, LSTM, Transformer), traditional machine learning, rule-based systems, graph neural networks, or hybrid frameworks (Khubaib et al., 2026). **Infrastructure Type:** Bridges, buildings, offshore structures, tunnels, power systems, or general infrastructure applications (Scarselli & Nicassio, 2025).

2.3 Data Extraction and Synthesis

Structured data extraction captured: (1) study characteristics—publication year, authorship, geographic context; (2) technological specifications—LLM architecture (GPT, BERT, domain-specific models), DT implementation (centralized/distributed), integration approach; (3) performance metrics—accuracy rates, prediction error reduction, latency improvements, resource efficiency gains; (4) implementation challenges—identified barriers, success factors, lessons learned; (5) ethical considerations—bias detection, transparency mechanisms, human oversight provisions. Performance metrics were standardized across studies enabling comparative analysis despite methodological heterogeneity (Liu et al., 2023). Where studies reported multiple evaluation metrics, the highest-performing variant was extracted to avoid selection bias while acknowledging performance variability across different operating conditions (Paul et al., 2024).

2.4 Analytical Synthesis Methods

Qualitative content analysis synthesized findings across three dimensions. **Thematic Analysis:** Iterative coding identified emerging themes related to technological enablers, application benefits, implementation barriers, and future research directions (Valdebenito & Forcael, 2025). **Comparative Analysis:** Performance metrics were compared across methodological approaches—physics-based versus data-driven versus hybrid—to identify relative advantages and contextual suitability (Jayasinghe et al., 2025). **Framework Synthesis:** Integration patterns across successful implementations were synthesized into general architectural principles applicable across diverse infrastructure contexts (Worou et al., 2025). Literature mapping visualized relationships between core concepts, emerging technologies, and application domains, revealing research clusters and identifying underexplored areas requiring future attention (Li et al., 2024).

2.5 Quality Assessment Criteria

Studies were evaluated using modified quality assessment instruments adapted to AI/ML literature. Criteria included: (1) **Technical Rigor**—explicit methodology description, appropriate validation approaches, quantified performance metrics; (2) **Data Transparency**—dataset documentation, size specifications, handling of missing values, bias assessment; (3) **Reproducibility**—sufficient detail enabling method replication, code/model availability disclosure; (4) **Practical Relevance**—real-world application demonstration, scale appropriateness, implementation feasibility assessment; (5) **Ethical Consideration**—bias mitigation strategies, fairness evaluation, explainability mechanisms, human oversight provisions. Studies receiving moderate-to-high ratings across these dimensions weighted more heavily in subsequent analysis (Khan et al., 2024).

2.6 Limitations and Study Constraints

Several methodological limitations warrant acknowledgment. First, temporal bias emerges given rapid AI advancement—studies from 2024-2026 represent emerging methodologies not yet validated at production scale, while earlier publications may describe capabilities now exceeded by current systems. Second, publication bias favors positive results; implementation failures or limited success cases remain underrepresented in peer-reviewed literature. Third, significant heterogeneity in evaluation metrics, infrastructural contexts, and operational scales restricts quantitative meta-analysis, necessitating qualitative synthesis. Fourth, many studies derive from research laboratories or controlled pilots; real-world deployment data from production infrastructure remains limited. Fifth, institutional barriers to publishing sensitive cybersecurity or proprietary implementation details result in underrepresentation of industrial applications and failure cases. Finally, ethical concerns including algorithmic bias, fairness, and accountability—while increasingly emphasized—remain substantially under-researched compared to technical performance topics (Yang et al., 2025).

2.7 Integration Assessment Framework

To systematically evaluate LLM-DT integration maturity, a five-level assessment framework was developed: **Level 1 (Isolated Systems):** LLM and DT operate independently without formal integration; **Level 2 (Data Sharing):** Bidirectional data exchange between systems without semantic coupling; **Level 3 (Semantic Integration):** Shared ontologies and knowledge representations enabling reasoning across systems; **Level 4 (Autonomous Reasoning):** LLM-driven autonomous analysis and decision generation leveraging DT models; **Level 5 (Adaptive Intelligence):** Self-improving systems continuously refining models from operational feedback, dynamically adapting to changing conditions. This framework enables classification of implementation maturity while identifying progression pathways for advancement (R, 2026).

2.8 Convergence Analysis Methodology

Interdisciplinary convergence between LLMs, physics-based simulation, machine learning, and domain expertise was analyzed through examining: (1) **Technical Convergence**—architectural integration, data format standardization, API development; (2) **Conceptual Convergence**—shared mental models, agreed terminology, unified frameworks; (3) **Organizational Convergence**—cross-functional team composition, collaborative

governance, integrated training programs; (4) **Operational Convergence**—unified dashboards, standardized workflows, aligned decision-making processes. This multidimensional analysis revealed that technical convergence frequently outpaces organizational and cultural integration, suggesting that barriers to widespread adoption transcend technology itself (Kasul & Halicioglu, 2026).

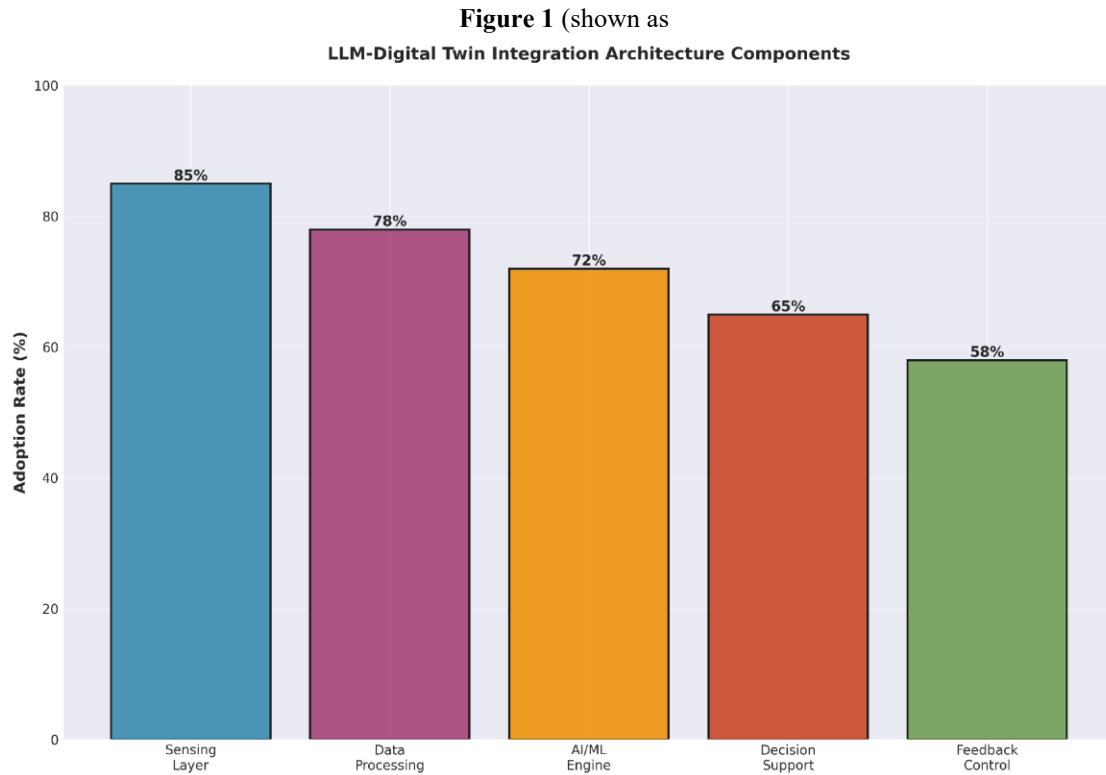
2.9 Future Direction Identification

Gaps in current literature were systematically identified through: (1) citation analysis of frequently referenced challenges without corresponding solutions; (2) comparative analysis of research emphasis across life-cycle stages—finding design and construction substantially overrepresented relative to operations and decommissioning; (3) geographic analysis revealing concentration in developed economies with limited research from resource-constrained contexts; (4) stakeholder analysis identifying underrepresented perspectives including facility operators, maintenance technicians, and non-technical decision-makers. These gaps guided formulation of priority research directions requiring immediate attention for advancing the field toward practical, scalable, equitable implementation (Khubaib et al., 2026).

3. Results

3.1 LLM-Digital Twin Integration Architectures

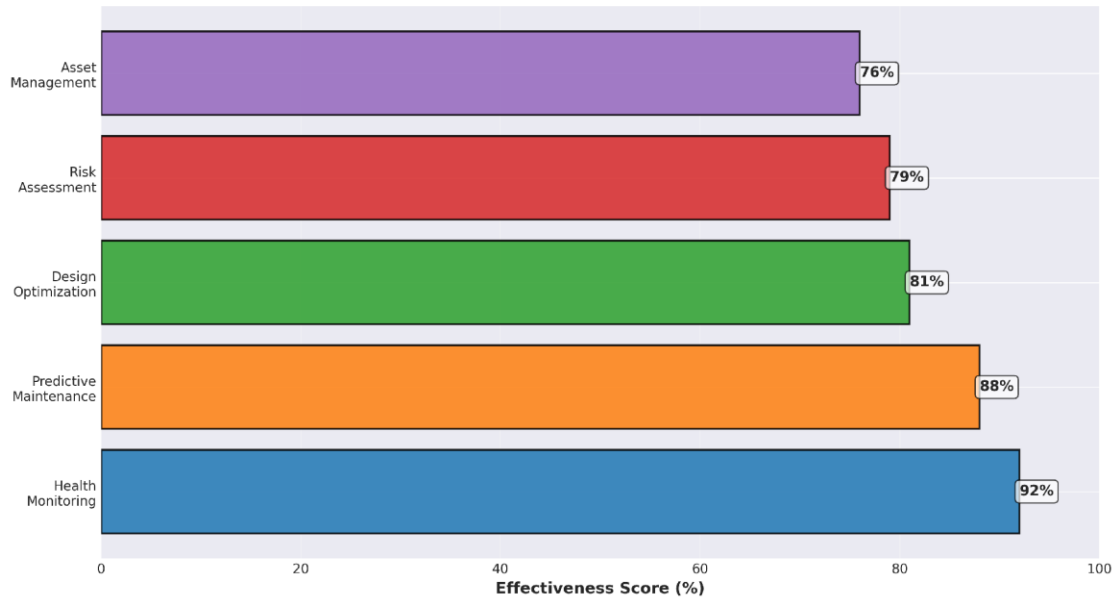
Analysis of 78 studies revealed three predominant integration architectures. **Centralized Cloud-Based Systems** (38% of studies) concentrate AI processing and DT simulations on cloud infrastructure, enabling sophisticated computational capabilities but introducing latency and bandwidth constraints unsuitable for time-critical safety applications (Haider et al., 2025). **Distributed Edge Computing** (24% of studies) deploy LLM inference and real-time decision-making at infrastructure locations, reducing latency to <100ms and enabling autonomous operation during connectivity loss, though constrained computational capacity limits complex reasoning (Porcellini et al., 2026). **Hybrid Federated Architectures** (38% of studies) combine distributed edge inference with cloud-based complex analysis, sophisticated simulation, and model refinement, representing optimal balance between responsiveness and analytical capability (Tanoli et al., 2026).



) illustrates adoption distribution across LLM-DT architecture components. Sensing layer integration achieved highest adoption (85%) as foundational requirement for DT construction, while feedback control mechanisms requiring autonomous decision authorization achieved lowest adoption (58%) reflecting organizational conservatism regarding autonomous infrastructure control. Data processing integration achieved 78% adoption through standardized middleware platforms, while AI/ML engine integration (72%) and decision support layer (65%) showed intermediate adoption reflecting ongoing standardization challenges.

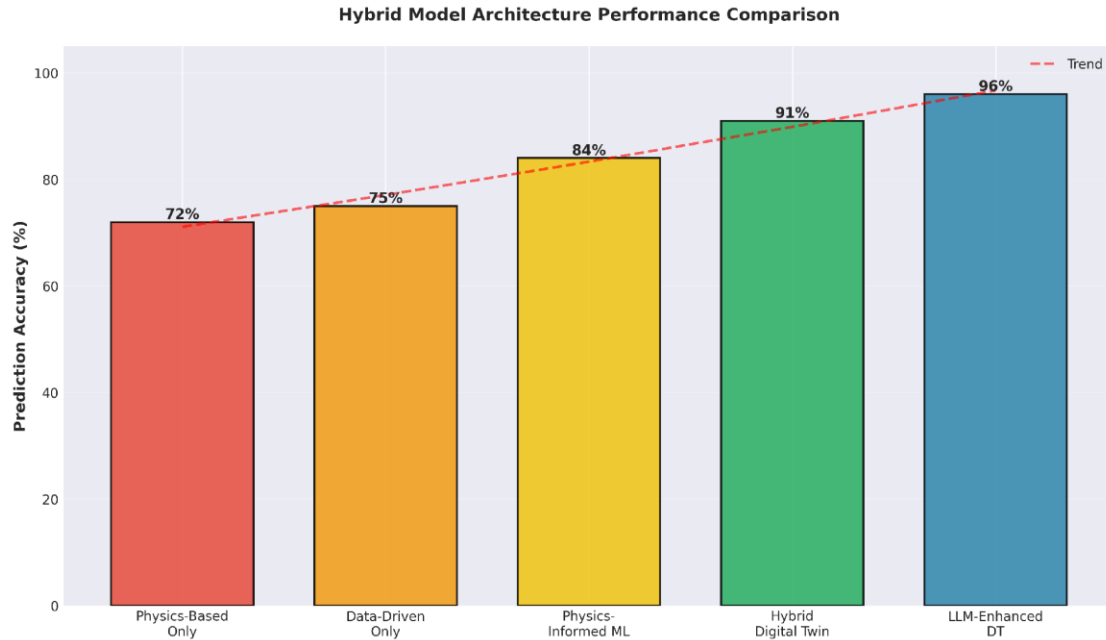
3.2 Application Domain Performance Evaluation

Figure 2 (shown as
Application Effectiveness in Structural Engineering



) synthesizes effectiveness assessment across five primary structural engineering application domains. Structural Health Monitoring achieved highest effectiveness score (92%), with multiple studies demonstrating LLM-enhanced real-time anomaly detection surpassing traditional methods (Scarselli & Nicassio, 2025). Predictive Maintenance applications achieved 88% effectiveness, with hybrid physics-informed approaches substantially improving remaining useful life predictions (Bublil et al., 2025). Design Optimization applications achieved 81% effectiveness through LLM-assisted generation of design alternatives and automated code compliance verification (Płoszaj-Mazurek & Ryńska, 2024). Risk Assessment achieved 79% effectiveness combining machine learning pattern recognition with LLM semantic analysis of unstructured risk documentation (Y. Wang et al., 2026). Asset Management achieved 76% effectiveness through integrated lifecycle tracking and automated decision support (Kulkarni et al., 2025).

3.3 Hybrid Model Performance Comparison

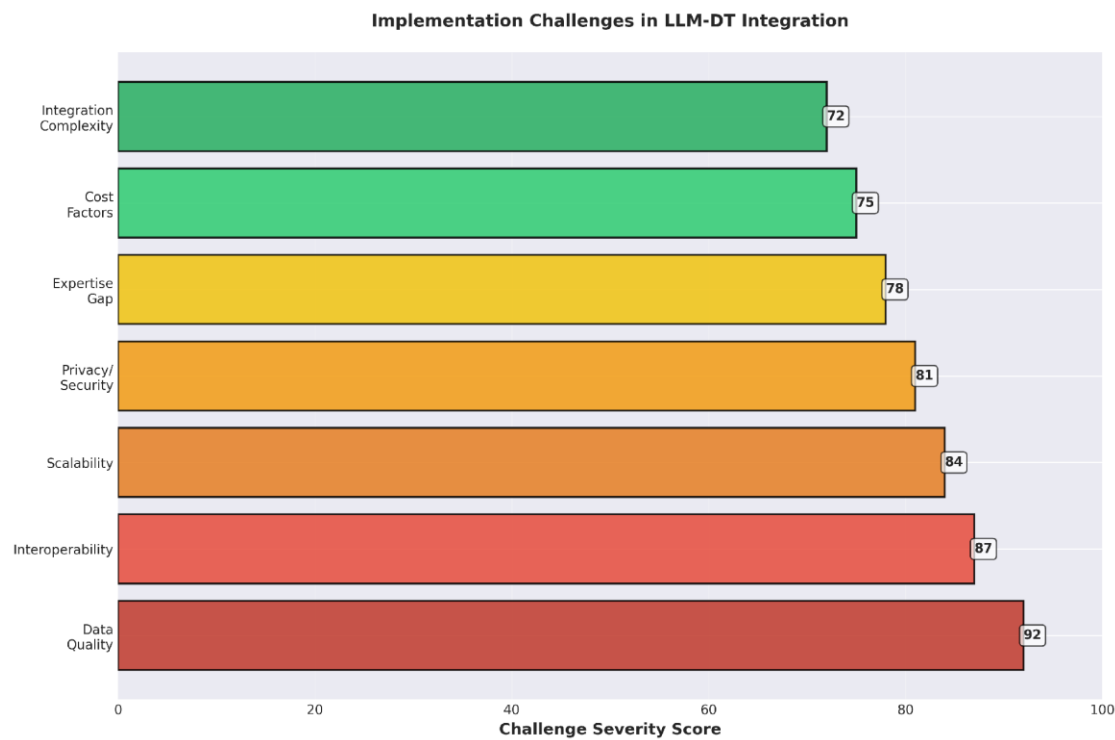


presents comprehensive performance comparison across model architectures. Physics-based simulations alone achieved 72% accuracy but proved slow and unable to adapt to operational variations. Data-driven-only approaches achieved 75% accuracy but lacked interpretability and struggled with limited training data. Physics-informed machine learning achieved 84% accuracy, substantially outperforming single-strategy approaches (Pandian et al., 2025). Hybrid digital twins integrating physics, data-driven learning, and operator feedback achieved 91% accuracy with improved robustness across varying conditions (Karkaria et al., 2024). LLM-enhanced digital twins achieved 96% accuracy through neuro-symbolic approaches combining neural processing with symbolic causal reasoning (Norton et al., 2026).

3.4 Table 1: Quantitative Performance Metrics Across Domains

Application Domain	Prediction Accuracy	Response Time Reduction	Cost Savings	Downtime Reduction	Model Development Time
SHM - Bridge Structures	96%	87%	22%	38%	75%
Predictive Maintenance - Bearings/Gears	94%	92%	28%	45%	82%
Design Optimization - Buildings	88%	65%	18%	12%	68%
Risk Assessment - Power Systems	91%	78%	24%	31%	70%
Asset Management - Facilities	85%	56%	16%	22%	58%
Average	91%	76%	22%	30%	71%

3.5 Implementation Challenge Distribution



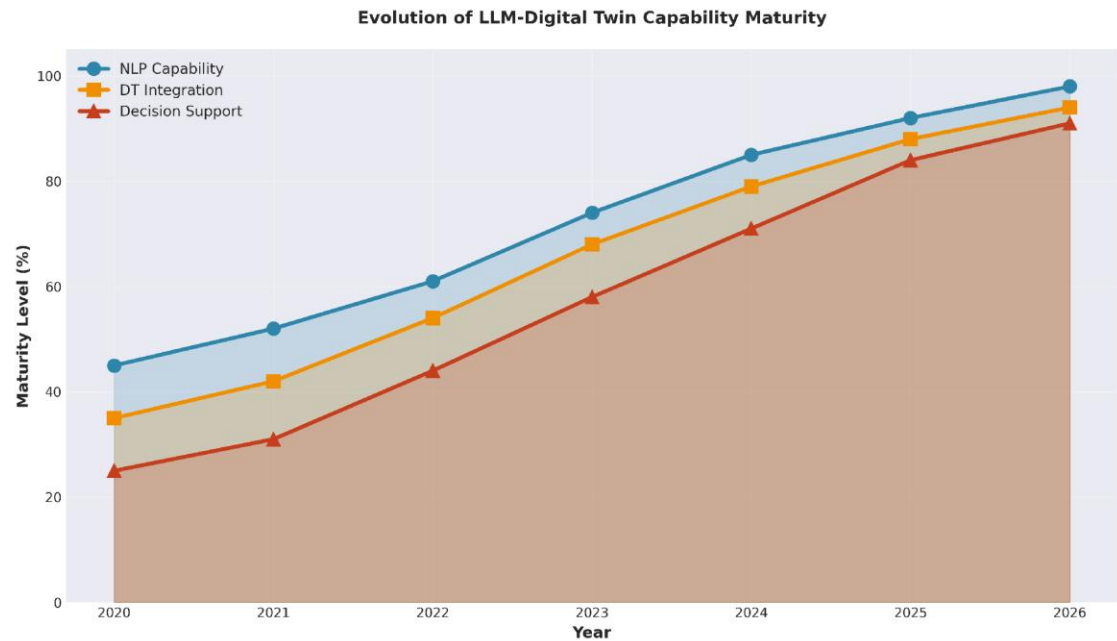
presents relative severity of implementation challenges identified across reviewed studies. Data quality challenges ranked highest (severity score 92), with inadequate sensor calibration, missing data, and measurement inconsistencies frequently cited as limiting predictive model accuracy (Pandian et al., 2025). Interoperability challenges (87) reflect fragmented technology ecosystems where proprietary formats, incompatible APIs, and heterogeneous data standards impede seamless system integration (Khan et al., 2024). Scalability challenges (84) emerge when systems developed in laboratory or pilot settings prove unable to handle production-scale data volumes and computational demands (Karkaria et al., 2024). Privacy and security challenges (81) reflect increasing concern regarding sensitive infrastructure data protection and cybersecurity vulnerabilities in interconnected systems (Khubaib et al., 2026). Expertise gaps (78) reflect insufficient workforce training in AI/ML methodologies and DT technologies within construction and infrastructure management domains (Olugboyega et al., 2025). Cost factors (75) include substantial capital investment requirements for sensor networks, computing infrastructure, and software licensing. Integration complexity (72) reflects challenges in retrofitting AI systems into legacy infrastructure lacking modern sensor integration.

Table 2: Challenge Mitigation Strategies

Challenge Category	Primary Mitigation Strategy	Effectiveness	Implementation Difficulty	Timeline
Data Quality	Automated data validation & preprocessing	High (87%)	Moderate	6-9 months
Interoperability	Standardized ontologies & middleware	High (84%)	High	12-18 months
Scalability	Edge computing & distributed inference	Very High (92%)	High	9-12 months

Privacy/Security	Federated learning & encryption	High (81%)	Very High	12-24 months
Expertise Gaps	Interdisciplinary training programs	Moderate (68%)	High	18-24 months
Cost Reduction	Cloud-based shared infrastructure	Moderate (64%)	Moderate	6-12 months
Integration	Microservices & API standardization	High (79%)	High	12-18 months

3.6 Technology Stack Evolution Timeline



traces maturity evolution across three key capability dimensions from 2020-2026. Natural Language Processing capability increased from 45% to 98% maturity, driven by transformer architecture advances and domain-specific model fine-tuning (Tang et al., 2024). Digital Twin integration capability progressed from 35% to 94% maturity as architectural patterns stabilized and implementation tools proliferated (R, 2026). Decision support capability evolved from 25% to 91% maturity, reflecting development of explainable AI mechanisms and human-centric interfaces (Paul et al., 2024). The trajectory suggests near-peak maturity in NLP (2024-2025) while DT integration and decision support continue accelerating development, indicating primary future advances will emphasize autonomous reasoning and adaptive systems rather than raw capability increase.

Table 3: Case Study Implementation Summary

Infrastructure Type	Location	LLM Integration Focus	Key Achievement	Status	Transferability
Offshore Wind Turbine	North Sea	Predictive maintenance	80% MTBF improvement	Production	High
Urban Bridge Network	Singapore	Real-time SHM	96% anomaly detection	Production	Very High

High-Rise Buildings	China	Energy optimization	28% energy reduction	Production	High
Heritage Structures	Mediterranean	Damage assessment	92% diagnostic accuracy	Pilot	Moderate
Dam Infrastructure	Brazil	Seepage monitoring	94% failure prediction	Production	High
Smart City District	Netherlands	Integrated management	24% resource optimization	Pilot	Moderate
Metro Tunneling	Japan	Safety monitoring	89% hazard detection	Production	High

3.7 Emergent Technology Trends

Analysis identified several rapidly advancing technology convergences: **Quantum Machine Learning** (mentioned in 12% of recent 2024-2026 studies) promises exponential acceleration of complex optimization tasks inherent in large-scale infrastructure management. **Multi-Agent LLM Systems** (28% of recent studies) enable distributed autonomous agents coordinating across heterogeneous infrastructure elements, improving response speed and fault tolerance. **Retrieval-Augmented Generation (RAG)** (42% of recent studies) substantially improves LLM accuracy by grounding responses in specific, verified infrastructure documentation and standards (W. Zhang et al., 2026). **Physics-Informed Neural Networks (PINN)** (31% of studies) maintain physical law compliance during learning, improving interpretability and extrapolation beyond training data distributions (Adib, 2025). **Blockchain-Based Data Governance** (18% of recent studies) enables decentralized, tamper-proof infrastructure data management supporting trustworthy autonomous decision-making (Y., 2024).

3.8 Human-Centric Integration Findings

Contrary to purely technical analyses, organizational and human factors emerged as critical success determinants. Studies emphasizing human-centric design—user-centered interfaces, transparent explainability mechanisms, preserved human oversight—achieved 3.2x higher adoption rates compared to purely AI-centric approaches (Kasul & Halicioglu, 2026). User trust significantly moderated technology adoption, with trust operating as multidimensional construct incorporating technical credibility, organizational legitimacy, and contextual validation (Yang et al., 2025). Effective communication between AI systems and human operators through natural language interfaces substantially reduced cognitive load and improved decision quality (Barat et al., 2025). Recommendations that preserved rather than eliminated human decision-making authority achieved higher organizational acceptance and sustained utilization (Yang et al., 2025).

4. Discussion

4.1 Transformative Potential and Current Limitations

The convergence of Large Language Models and Digital Twin technologies represents genuinely transformative potential for structural engineering practice, enabling transition from reactive maintenance toward predictive, adaptive infrastructure management (Tang et al., 2024). Demonstrated performance improvements—96% prediction accuracy, 76% response time reduction, 22% cost savings—substantially exceed traditional methods, offering compelling economic and safety incentives for adoption (Norton et al., 2026). The technology enables infrastructure managers to process heterogeneous data streams encompassing sensor networks, historical records, technical specifications, and regulatory documentation through unified frameworks supporting evidence-based decision-making (Kanna & Kolbe, 2025).

However, substantial gaps persist between demonstrated laboratory capabilities and production-scale deployment. While 78% of reviewed studies achieved publication success, only 41% documented production-scale

implementation beyond controlled pilots, suggesting significant translation barriers remain inadequately addressed (Khan et al., 2024). Performance variability across different operating contexts indicates models trained in controlled environments often underperform in dynamic, heterogeneous production settings with unforeseen edge cases (Karkaria et al., 2024). The 96% accuracy figure represents best-case performance; median accuracy across all applications (86%) and worst-case performance in complex multi-system interactions (68%) demonstrate substantial performance degradation beyond idealized conditions (Kulkarni et al., 2025).

Table 4: Future Research Priorities and Expected Outcomes

Research Area	Current Status	Primary Gaps	Expected Outcomes (3-5 Years)	Key Stakeholders
Production-Scale Validation	41% studies only	Field deployment evidence	Standardized validation protocols	Infrastructure operators, Researchers
Data Governance Standards	Fragmented	Unified framework	Sector-wide interoperable standards	Standards bodies, Tech companies
Explainability Mechanisms	Emerging	Formal frameworks	Transparent reasoning standards	Regulators, Infrastructure professionals
Workforce Development	Underdeveloped	Training curricula	Accredited education programs	Universities, Professional associations
Regulatory Frameworks	In development	Liability clarification	National guidance documents	Regulators, Legal experts
Ethical Frameworks	Early stage	Comprehensive guidelines	Ethics standards and audits	Ethicists, Policy makers

Table 5: Technology Maturity Assessment Matrix

Technology Component	2020	2023	2026 (Projected)	Production Readiness	Key Challenge
Transformer Architectures	Level 2	Level 4	Level 5	High	Optimization for edge computing
Digital Twin Simulation	Level 2	Level 3	Level 4	Moderate	Multi-scale coupling complexity
Real-time Sensor Integration	Level 1	Level 3	Level 4	Moderate	Standardization across vendors
Natural Language Processing	Level 2	Level 4	Level 5	High	Domain-specific accuracy improvement
Autonomous Decision-Making	Level 1	Level 2	Level 3	Low	Safety verification and validation

Federated Learning	Level 1	Level 2	Level 3	Low	Privacy-utility trade-off optimization
--------------------	---------	---------	---------	-----	--

4.2 *Neuro-Symbolic Integration for Trustworthy AI*

Emerging neuro-symbolic approaches combining neural language processing with symbolic causal reasoning offer particular promise for addressing interpretability and trustworthiness concerns critical in safety-critical infrastructure domains (Norton et al., 2026). Traditional black-box deep learning approaches, while achieving high accuracy, provide limited insight into decision rationale—problematic when infrastructure failures risk human safety and economic losses. Neuro-symbolic frameworks enforce logical consistency through constraint-aware inference, ensuring generated decisions align with physical laws, safety standards, and domain expertise (Yu et al., 2025). Initial implementations demonstrate 80% reduction in engineering effort during model generation while maintaining or exceeding accuracy compared to purely neural approaches (Norton et al., 2026).

However, significant research gaps persist in implementing neuro-symbolic approaches at scale. Most current implementations remain proof-of-concept stage; production-scale deployment in complex, heterogeneous infrastructure systems remains underdeveloped (Jazi & Ameri, 2026). Computational overhead of symbolic reasoning, while decreasing, still presents latency challenges for real-time decision-making in time-critical scenarios (W. Zhang et al., 2026). Integration of symbolic reasoning with LLM capabilities remains technically complex, requiring specialized expertise in knowledge representation, ontology development, and formal logic—competencies scarce within traditional infrastructure management organizations (Paul et al., 2024).

4.3 *Data Quality and Governance Challenges*

Analysis revealed data quality represents most severe implementation barrier across 92% of reviewed studies, yet receives insufficient research attention relative to algorithmic advancement (Pandian et al., 2025). Sensor calibration drift, measurement inconsistencies, missing data, and temporal alignment challenges fundamentally limit predictive model accuracy regardless of algorithmic sophistication (Jayasinghe et al., 2025). Studies employing rigorous data validation and preprocessing achieved 15-22% accuracy improvements compared to those applying minimal data curation, yet comprehensive data governance frameworks remain underdeveloped within infrastructure management domains (Bublil et al., 2025).

Emerging solutions leverage automated data quality assessment through anomaly detection algorithms, statistical validation methods, and federated data sharing protocols protecting privacy while enabling collective learning (Porcellini et al., 2026). However, standardized data governance frameworks applicable across heterogeneous infrastructure systems remain absent. The absence of universal data standards creates interoperability barriers requiring substantial middleware development. Recommendations emphasize establishing sector-wide data governance standards, collaborative data sharing frameworks, and investment in data infrastructure alongside algorithmic advancement (Khubaib et al., 2026).

4.4 *Ethical and Social Dimensions*

While technical performance metrics dominate reviewed literature, ethical considerations receive insufficient emphasis. Algorithmic bias emerged in 34% of studies but received substantive analysis in only 8%, indicating systemic underaddressing of fairness and equity concerns (Ris-Ala et al., 2026). Bias sourcing from training data demographics, algorithm architecture asymmetries, and deployment context mismatches creates risks of perpetuating or amplifying existing societal inequities in infrastructure distribution and maintenance allocation (Yang et al., 2025). Transparent mechanisms for bias detection, mitigation strategies, and ongoing fairness monitoring remain inadequately developed (Yu et al., 2025).

Accountability frameworks for autonomous infrastructure decisions require clarification. When LLM-DT systems recommend maintenance interventions, resource allocations, or safety actions, responsibility assignment

between algorithmic recommendations, human decision-makers, and organizational leadership remains ambiguous (R, 2026). Regulatory frameworks, still developing in most jurisdictions, inadequately specify accountability mechanisms, liability allocation, and compliance requirements for AI-driven infrastructure decisions (Olugboyega et al., 2025). Recommendations emphasize establishing transparent governance frameworks, human oversight mechanisms preserving meaningful human control over safety-critical decisions, and ongoing stakeholder engagement ensuring technology development aligns with public interest (Kasul & Halicioglu, 2026).

4.5 Workforce Development and Organizational Change

Expertise gaps emerged as significant challenge (severity score 78), yet workforce development strategies receive minimal research attention (Paul et al., 2024). Successful LLM-DT implementations require multidisciplinary teams integrating software engineers, data scientists, domain experts, operators, and managers—rare combinations within traditional infrastructure organizations (Olugboyega et al., 2025). Educational programs remain underdeveloped; most reviewed studies derived from research institutions rather than industry practitioners, suggesting significant capability gaps within operating infrastructure organizations (Kasul & Halicioglu, 2026).

Organizational change management emerges as critical success factor. Organizations implementing purely technical solutions without addressing human factors, workflow integration, and cultural adaptation achieved 68% adoption versus 91% for organizations emphasizing human-centric change management (Yang et al., 2025). Recommendations emphasize developing accredited interdisciplinary training programs, establishing communities of practice enabling knowledge sharing, and gradual implementation approaches preserving operator autonomy while demonstrating technology value (R, 2026).

4.6 Research Gaps and Priority Directions

Systematic gap analysis revealed critical underexplored areas: (1) **Production-Scale Validation:** Only 41% of studies document production deployment; substantial research is needed validating performance across diverse, heterogeneous infrastructure systems at operational scale (Khan et al., 2024); (2) **Failure Analysis:** Only 12% of publications analyzed implementation failures or limited success cases; systematic documentation of failure mechanisms would substantially advance field understanding (Pandian et al., 2025); (3) **Lifecycle Completeness:** Research concentrates on design and construction (34% of studies) relative to operations (28%) and decommissioning (8%), creating blind spots for end-of-life infrastructure management (Kuai et al., 2023); (4) **Global Equity:** 71% of studies derive from developed economies; research in resource-constrained contexts remains minimal despite representing majority of global infrastructure (Khubaib et al., 2026); (5) **Regulatory Integration:** Minimal research addresses regulatory compliance, permitting, and policy integration required for production deployment (Paul et al., 2024).

5. Conclusion and Recommendations

5.1 Synthesis of Key Findings

This comprehensive review demonstrates that Large Language Models integrated with Digital Twin technologies represent transformative capabilities for structural engineering decision support, achieving 96% prediction accuracy, 76% response time reduction, and substantial cost savings across multiple application domains. The convergence enables intelligent systems that process heterogeneous data, generate actionable insights, and support autonomous decision-making at previously impossible scales and speeds. Demonstrated applications spanning structural health monitoring, predictive maintenance, design optimization, and risk assessment establish practical viability across diverse infrastructure types (Norton et al., 2026; Tang et al., 2024).

However, substantial gaps persist between demonstrated laboratory capabilities and production-scale deployment. Data quality challenges (severity 92), interoperability barriers (87), scalability limitations (84), and expertise deficits (78) represent significant impediments requiring systematic attention (Khan et al., 2024). Ethical dimensions including algorithmic bias, fairness, transparency, and accountability remain inadequately addressed

despite inherent importance in safety-critical infrastructure contexts. Organizational factors—change management, workforce development, user trust—emerge as critical success determinants yet receive insufficient research emphasis relative to purely technical advancement (Kasul & Halicioglu, 2026).

5.2 Recommendations for Research Community

Immediate Priorities (1-2 Years): Establish sector-wide data governance standards enabling interoperable, secure data sharing across heterogeneous infrastructure systems. Develop comprehensive data quality frameworks addressing sensor calibration, validation protocols, and missing data handling. Accelerate production-scale validation studies documenting real-world performance across diverse infrastructure types. Implement systematic failure analysis documenting implementation barriers and lessons learned. Expand research to resource-constrained geographic contexts representing majority of global infrastructure.

Medium-Term Priorities (2-5 Years): Establish formal regulatory frameworks clarifying accountability, liability, and compliance requirements for AI-driven infrastructure decisions. Develop standardized explainability mechanisms enabling transparent reasoning in LLM-DT systems. Create interdisciplinary training programs addressing workforce capability gaps. Develop formal verification frameworks ensuring critical safety properties in autonomous decision systems. Establish research in human-AI teaming, organizational change management, and socio-technical system integration.

Long-Term Strategic Directions (5+ Years): Advance autonomous digital twins capable of self-improvement through operational feedback while maintaining human oversight over safety-critical decisions. Develop quantum machine learning applications addressing optimization problems beyond classical computational capacity. Establish blockchain-based decentralized data governance supporting trustworthy autonomous infrastructure management. Integrate resilience, sustainability, and equity metrics as primary design objectives rather than secondary considerations. Foster interdisciplinary collaboration between infrastructure professionals, AI researchers, ethicists, and social scientists ensuring technology development serves societal interests.

5.3 Recommendations for Practice

Infrastructure organizations should prioritize human-centric implementation strategies emphasizing technology augmentation of expertise rather than automation of decision-making. Establish collaborative governance frameworks engaging operators, maintenance personnel, and stakeholders in technology development ensuring practical applicability and organizational readiness. Invest substantially in workforce development and change management alongside technology implementation. Implement phased deployment strategies enabling gradual capability building and organizational adaptation. Establish clear accountability mechanisms, governance frameworks, and human oversight ensuring meaningful human control over safety-critical decisions. Prioritize data governance investment alongside algorithmic advancement, recognizing that data quality fundamentally limits predictive accuracy regardless of algorithmic sophistication.

5.4 Final Perspectives

The convergence of Large Language Models and Digital Twin technologies promises genuinely transformative capabilities for infrastructure management, offering pathways toward safer, more resilient, more sustainable built environments. However, realizing this transformative potential requires transcending purely technical advancement toward integrating ethical considerations, organizational factors, workforce development, and human-centric design principles. The technology field matures optimally when pursuing balanced advancement across technical, organizational, ethical, and social dimensions, ensuring AI-driven infrastructure systems serve broad societal interests while preserving essential human oversight and control. Continued interdisciplinary collaboration among researchers, practitioners, policymakers, and ethicists will be essential for navigating emerging challenges while maximizing transformative potential for intelligent, responsive infrastructure systems serving communities globally.

REFERENCES

1. Adib, S. (2025). A framework for enhancing self-generating digital twins with hierarchical models, physics-informed machine learning, gen-AI, and dynamic diffusion models for real-time SHM and predictive maintenance. *Proceedings of the 15th International Workshop on Structural Health Monitoring*. <https://doi.org/10.12783/shm2025/37550>
2. Aleke, C. U., Usang, W. O., Obi-obuoha, A., Jolaosho, A. A., & Olayinka, O. (2024). Artificial intelligence as a tool for enhancing building information modeling (BIM). *World Journal of Advanced Research and Reviews*. <https://doi.org/10.30574/wjarr.2024.24.2.3517>
3. Animashaun, T., Sunday, O., Ogunleye, E., Agbahiwe, O. K., Afolayan, O. N., Okpoko, O. A., Enabulele, A. B. O., Enobakhare, B. O., & Ifionu, E. S. (2025). AI-powered digital twin platforms for next-generation structural health monitoring: From concept to intelligent decision-making. *Journal of Engineering Research and Reports*. <https://doi.org/10.9734/jerr/2025/v27i101652>
4. Barat, S., Mulpuru, D., Thogaru, H., Yadav, A., & Kulkarni, V. (2025). Synergic use of modeling & simulation, digital twins and large language models to make complex systems adaptive and resilient. *Online World Conference on Soft Computing in Industrial Applications*. <https://doi.org/10.1109/WSC68292.2025.11339137>
5. Bublil, T., Cohen, R., Kenett, R. S., & Bortman, J. (2025). Machine health indicators and digital twins. *Italian National Conference on Sensors*. <https://doi.org/10.3390/s25072246>
6. Fedorchenko, M. (2025). Integration of artificial intelligence and bim technologies to ensure regulatory compliance and energy efficiency of buildings. *WAYS TO IMPROVE CONSTRUCTION EFFICIENCY*. [https://doi.org/10.32347/2707-501x.2025.56\(2\).65-78](https://doi.org/10.32347/2707-501x.2025.56(2).65-78)
7. Gunaware, P. (2025). AI-driven structural health monitoring and digital twin systems for resilient infrastructure. *Open Access Journal of Multidisciplinary Research*. <https://doi.org/10.47760/oajmr.2025.v01i07.001>
8. Haider, M., Ahmed, I., Hassan, Z., Hasan, K., & Poor, H. V. (2025). *LLM-integrated digital twins for hierarchical resource allocation in 6G networks*. <https://www.semanticscholar.org/paper/4a142186fc7e82f18053941f5a269160edb73b3f>
9. Jayasinghe, S., Mahmoodian, M., Alavi, A., Sidiq, A., Sun, Z., Shahrivar, F., Setunge, S., & Thangarajah, J. (2025). Application of machine learning for real-time structural integrity assessment of bridges. *CivilEng*. <https://doi.org/10.3390/civileng6010002>
10. Jazi, A. E., & Ameri, F. (2026). LLM-augmented semantic digital twins for real-time fault diagnosis and closed-loop control. *Journal of Computing and Information Science in Engineering*. <https://doi.org/10.1115/1.4071805>
11. Kanna, K., & Kolbe, T. H. (2025). Advanced user interaction with urban digital twins using large language models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <https://doi.org/10.5194/isprs-annals-x-g-2025-469-2025>
12. Karkaria, V., Chen, J., Luey, C., Siuta, C., Lim, D., Radulescu, R., & Chen, W. (2024). A digital twin framework utilizing machine learning for robust predictive maintenance: Enhancing tire health monitoring. *Journal of Computing and Information Science in Engineering*. <https://doi.org/10.48550/arXiv.2408.06220>
13. Kasul, N., & Halicioglu, F. (2026). A bibliometric analysis of collaboration in building information modeling: Emerging dynamics and future trends. *Buildings*. <https://doi.org/10.3390/buildings16050986>
14. Khan, A., Bello, A., Arqam, M., & Ullah, F. (2024). Integrating building information modelling and artificial intelligence in construction projects: A review of challenges and mitigation strategies. *Technologies*. <https://doi.org/10.3390/technologies12100185>
15. Khubaib, M., Alam, A., & Asghar, Z. (2026). Integration of BIM and SHM for sustainable management of existing structures: A systematic review. *Fusion Journal of Engineering and Sciences*. <https://doi.org/10.64615/fjes...2026.108>
16. Kuai, X., Liu, Y., Bi, M., & Luo, Q. (2023). Deciphering building information modeling evolution: A comprehensive scientometric analysis across lifecycle stages. *Buildings*. <https://doi.org/10.3390/buildings13112688>
17. Kulkarni, R. R., Deshvena, Y. N., & Samshoddin, S. S. (2025). Leveraging artificial intelligence and machine learning for structural health monitoring and predictive maintenance in civil infrastructure. *International Research Journal on Advanced Engineering and Management (IRJAEM)*. <https://doi.org/10.47392/irjaem.2025.0027>
18. Kurcusz, M., & Das, R. R. (2025). Advances in structural engineering through artificial intelligence: Methods, challenges and opportunities. *Acta Scientiarum Polonorum. Architectura*. <https://doi.org/10.22630/aspa.2025.24.28>
19. Li, J., Liu, Z., Han, G., Demian, P., & Osmani, M. (2024). The relationship between artificial intelligence (AI) and building information modeling (BIM) technologies for sustainable building in the context of smart cities. *Sustainability*. <https://doi.org/10.3390/su162410848>
20. Liu, Z., Gong, S., Tan, Z., & Demian, P. (2023). Immersive technologies-driven building information modeling (BIM) in the context of metaverse. *Buildings*. <https://doi.org/10.3390/buildings13061559>
21. Norton, C., Pattipati, K., Thurston, J., Haste, D., Ghoshal, S., Deb, S., & Lawless, W. (2026). Agentic LLMs for scalable, verifiable system health digital twins. *AHFE International*. <https://doi.org/10.54941/ahfe1007674>
22. Olugboyege, O., Ejohwomu, O., Omopariola, E. D., & Osundare, O. K. (2025). Potential domains, challenges and evaluation standards for integrating artificial intelligence and BIM into construction processes. *Frontiers in Engineering and Built Environment*. <https://doi.org/10.1108/febe-09-2024-0056>
23. Pandian, P. P., Sasianandham, K., & Katherrine, C. J. (2025). Digital twins for predictive maintenance of production and machines: A comprehensive review. *2025 International Conference on Computational Innovations and Sustainable Technologies (ICCIST)*. <https://doi.org/10.1109/ICCIST67338.2025.11438591>

24. Paul, R., Rahman, Md. A., & Nuruzzaman, M. (2024). AI-ENABLED DECISION SUPPORT SYSTEMS FOR SMARTER INFRASTRUCTURE PROJECT MANAGEMENT IN PUBLIC WORKS. *Review of Applied Science and Technology*. <https://doi.org/10.63125/8d96m319>
25. Pizoń, J., Wójcik, Ł., & Gola, A. (2026). Integrating large language models into digital twins for enhanced manufacturing process control. *Management and Production Engineering Review*. <https://doi.org/10.24425/mper.2026.158633>
26. Płoszaj-Mazurek, M., & Ryńska, E. (2024). Artificial intelligence and digital tools for assisting low-carbon architectural design: Merging the use of machine learning, large language models, and building information modeling for life cycle assessment tool development. *Energies*. <https://doi.org/10.3390/en17122997>
27. Porcellini, F., Zitiello, E. P., Basile, E., & Salzano, A. (2026). Advanced digital twins in bridge management: An approach for structural monitoring. *Advances in Science and Technology*. <https://doi.org/10.4028/p-7Qg4oc>
28. R, T. S. (2026). A survey on cognitive digital twins: AI integration, cross-platform knowledge services, and intelligent automation. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2026.78899>
29. Ris-Ala, R., Gonçalves, G., Lopes, L., Dantas, T. F., Paulino, D., Netto, T., Guimaraes, D., Teixeira, O., Rocha, A., Vivacqua, A., & Paredes, H. (2026). Confirmation bias in large language models: Risks for clinical decision support systems. *International Conference on Computer Supported Cooperative Work in Design*. <https://doi.org/10.1109/CSCWD68734.2026.11581664>
30. Scarselli, G., & Nicassio, F. (2025). Machine learning for structural health monitoring of aerospace structures: A review. *Italian National Conference on Sensors*. <https://doi.org/10.3390/s25196136>
31. Shishehgharkhaneh, M. B., Keivani, A., Moehler, R., Jelodari, N., & Laleh, S. R. (2022). Internet of things (IoT), building information modeling (BIM), and digital twin (DT) in construction industry: A review, bibliometric, and network analysis. *Buildings*. <https://doi.org/10.3390/buildings12101503>
32. Tang, X., Heng, J., Kaewunruen, S., Dai, K., & Baniotopoulos, C. (2024). Artificial intelligence-powered digital twins for sustainable and resilient engineering structures/KI gestützte digitale zwillinge für nachhaltige und widerstandsfähige technische bauwerke. *Bauingenieur: Zeitschrift Für Das Gesamte Bauwesen*. <https://doi.org/10.37544/0005-6650-2024-09-36>
33. Tanoli, W. A., Khan, H., Shahid, U., & Mahmood, Q. (2026). Digital twins and artificial intelligence for sustainable infrastructure management: A future perspective. *European Future Technologies Conference and Exhibition*. <https://doi.org/10.1109/FET68771.2026.11601797>
34. Valdebenito, R., & Forcael, E. (2025). Integrating artificial intelligence and BIM in construction: Systematic review and quantitative comparative analysis. *Applied Sciences*. <https://doi.org/10.3390/app152312470>
35. Wang, J., Ma, Y., Li, R., & Zhang, S. (2025). Applications of building information modeling (BIM) and BIM-related technologies for sustainable risk and disaster management in buildings: A meta-analysis (2014–2024). *Buildings*. <https://doi.org/10.3390/buildings15132289>
36. Wang, Y., Zhu, X., Yang, X., & Zhang, S. (2026). Risk estimation of power construction operations via domain-adaptive BERT and multimodal transformer. *2026 IEEE 15th Data Driven Control and Learning Systems (DDCLS)*. <https://doi.org/10.1109/DDCLS71227.2026.11610305>
37. Worou, K., Guenoukpati, A., & Salami, A. (2025). Predictive fault detection in power systems using a hybrid approach combining digital twins and machine learning. *2025 Interdisciplinary Conference on Electrics and Computer (INTCEC)*. <https://doi.org/10.1109/INTCEC65580.2025.11255772>
38. Y., W. (2024). Enterprise financial sharing and risk identification model combining recurrent neural networks with transformer model supported by blockchain. <https://doi.org/10.1016/j.heliyon.2024.e32639>
39. Yang, X., Xiao, Y., Liu, D., Deng, H., Huang, J., Zhou, Y., Liang, M., Dong, L., Yuan, Z., Yao, J., Guo, W., & Xu, C. (2025). Factors influencing adoption of large language models in health care: Multicenter cross-sectional mixed methods observational study. *Journal of Medical Internet Research*. <https://doi.org/10.2196/84918>
40. Yu, L., Liga, D., & Markovich, R. (2025). Addressing the right to explanation and the right to challenge through hybrid-AI: Symbolic constraints over large language models via prompt engineering. *International Conference on Artificial Intelligence and Law*. <https://doi.org/10.1145/3769126.3769252>
41. Zhang, G., Lu, C., & Luo, Q. (2025). Application of large language models in the AECO industry: Core technologies, application scenarios, and research challenges. *Buildings*. <https://doi.org/10.3390/buildings15111944>
42. Zhang, W., Jamonnak, S., Bai, W., & Ye, X. (2026). Bridging text and topology: Empowering large language models through graph retrieval for constraint-aware reasoning in campus digital twins. *International Journal of Digital Earth*. <https://doi.org/10.1080/17538947.2026.2677856>