DOI: 10.36297/vw.jei.v7i4.101

VW Engineering International, Volume: 7, Issue: 4, 1-18

Artificial Intelligence and Machine Learning in Mechanical Engineering: A Scientific Literature Review on Emerging Paradigms and Industrial Transformation

Md Irfanul Haque Siddiqui [1,*]

[1]Department of Mechanical Engineering, King Saud University, Riyadh 11451, Saudi Arabia
*Corresponding author's email: msiddiqui2.c@ksu.edu.sa

Received: Aug. 22, 2025
Accepted: Sep. 28, 2025
Published online: Oct. 03, 2025

Abstract: Artificial Intelligence (AI) and Machine Learning (ML) represent a transformative paradigm shift in mechanical engineering (ME), moving the discipline beyond traditional analytical and computational methodologies toward data-driven and physics-informed intelligence. This comprehensive scientific literature review critically synthesizes the state-of-the-art applications of AI/ML across the core domains of ME, including design optimization, smart manufacturing, prognostic health management (PHM), dynamic system control, and computational mechanics. The analysis details how advanced techniques, such as hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) models, overcome the challenges of predictive maintenance by effectively analyzing temporal sensor data, often achieving superior performance metrics, such as Mean Absolute Percentage Error (MAPE) below 0.36%. Simultaneously, Physics-Informed Neural Networks (PINNs) revolutionize simulation speed and physical consistency in complex phenomena like fracture and fluid dynamics, providing computationally reliable surrogate models, sometimes offering speedups of four orders of magnitude in online computation. Furthermore, the review examines the crucial shift toward inverse design methodologies and the utilization of robust hybrid architectures, such as Evolutionary Reinforcement Learning (ERL), which enhances the stability and safety of autonomous control systems by mitigating issues like brittle convergence. A critical discussion addresses major integration hurdles, including model interpretability (Explainable AI, or XAI) due to the complexity of large-scale models, inherent data scarcity in industrial settings, and the necessity for establishing regulatory frameworks to govern AI deployment in safety-critical mechanical systems. The review concludes by underscoring the imperative for responsible innovation and deep interdisciplinary collaboration to responsibly harness the full potential of AI/ML for sustainable and advanced mechanical systems.

Keywords: Artificial Intelligence, Machine Learning, Mechanical Engineering, Physics-Informed Neural Networks, Smart Manufacturing, Prognostic Health Management

1. Introduction

Mechanical engineering (ME) serves as a foundational discipline for global infrastructure, energy production, manufacturing, and transportation systems. Historically, advancements in ME have been driven by empirical experimentation, followed by the adoption of analytical modeling, and later, powerful computational tools such as Finite Element Analysis (FEA) and Computational Fluid Dynamics (CFD). The current era, often termed the Fourth Industrial Revolution (4IR), marks a profound convergence of physical and digital technologies, wherein data, connectivity, and computational intelligence fundamentally reshape the engineering workflow. This convergence relies heavily on the capabilities of Artificial Intelligence (AI) and Machine Learning (ML) to process the massive volumes of data generated by the Internet of Things (IoT) sensors, manufacturing systems, and high-fidelity simulations. The adoption of AI/ML enables mechanical engineers to tackle problems previously deemed computationally or analytically intractable [1]. These methods facilitate the transition from optimizing an existing design or process to generating entirely new solutions based purely on desired performance criteria paradigm known as inverse design [11]. This transformation is not merely an improvement in tool speed; it represents a fundamental change in how engineering knowledge is acquired, represented, and utilized, pushing the boundaries of innovation toward smarter, more efficient, and more sustainable systems.

While the recent growth of AI in engineering appears sudden, ML techniques have been applied in specific fields of materials science and engineering since the early 1960s [18]. However, the current explosive growth is attributed to three synergistic factors: the widespread availability of high-performance computing resources (particularly Graphics Processing Units, or GPUs), the generation of massive, high-quality datasets, and significant algorithmic breakthroughs, primarily in Deep Learning (DL). DL architectures, characterized by multiple layers of interconnected nodes, can automatically learn hierarchical features directly from raw data, overcoming the reliance on manual feature engineering that limited earlier ML approaches. This maturation has allowed AI to permeate critical engineering functions. For instance, the Materials Genome Initiative (MGI) has been significantly accelerated by AI systems that drive optimization and discovery by analyzing large datasets [17]–[19]. Similarly, the adoption of DL in manufacturing has transitioned quality control from reactive inspection to real-time, adaptive process optimization [3]. The synthesis presented in this review demonstrates that AI/ML is moving beyond specialized applications to become an indispensable, generalized toolkit for the modern mechanical engineer. The rapid growth of this field is evidenced by a major review identifying over 14,000 related publications since 2016 [20].

For the purposes of this review, it is essential to clearly delineate the core concepts of AI, ML, and DL, and to introduce the key algorithmic families relevant to mechanical systems:

1. **Artificial Intelligence (AI):** The broadest concept, referring to the capability of a machine to imitate intelligent human behavior, encompassing reasoning, problem-solving, planning, and learning.

2. **Machine Learning (ML):** A subset of AI where systems learn from data, identify patterns, and make decisions with minimal explicit programming.

3. **Deep Learning (DL):** A subset of ML utilizing deep neural networks (DNNs) with multiple hidden layers, capable of modeling highly complex, nonlinear relationships and automatically extracting sophisticated features from unstructured data (e.g., sensor signals, images, or simulation results).

Key algorithmic families applied in ME include:

- **Supervised Learning:** Training models (such as Convolutional Neural Networks, CNNs, and Recurrent Neural Networks, RNNs/LSTMs) on labeled data to predict outcomes (classification or regression), widely used for Remaining Useful Life (RUL) estimation and material property prediction.

- **Unsupervised Learning:** Identifying hidden patterns or clusters in unlabeled data, crucial for real-time anomaly detection and condition monitoring.

- **Reinforcement Learning (RL):** Training agents to make sequential decisions in an environment to maximize a cumulative reward, essential for dynamic system control, production scheduling [21], and robotics [22].

- **Physics-Informed Neural Networks (PINNs):** A hybrid approach where the architecture or loss function of a neural network is constrained by governing physical equations (e.g., Partial Differential Equations, PDEs), ensuring solutions are physically consistent, particularly useful in computational mechanics [6].

The following table 1 organizes these paradigms with their primary functions within mechanical engineering domains:

Table 1: AI/ML Paradigms and Core Mechanical Engineering Applications

AI/ML Paradigm	Core Mechanical Engineering Application	Primary Function/Benefit
Supervised Learning (e.g., CNN, LSTM)	Remaining Useful Life (RUL) Estimation, Material Property Prediction, Quality Inspection	Prediction of failure time/properties, classification of defects.
Unsupervised Learning (e.g., Anomaly Detection)	Real-time Condition Monitoring, Process State Identification	Discovering hidden patterns, flagging novel anomalous behaviors.
Reinforcement Learning (RL)	Dynamic System Control, Robotic Path Planning 22, Production Scheduling 21	Real-time optimization, sequential decision-making in dynamic environments.
Physics-Informed Neural Networks (PINNs)	Computational Fluid Dynamics (CFD), Structural Mechanics (PDE Solving)	Creating fast, physically consistent surrogate models, inverse problem solving.6

This literature review provides a rigorous analysis of peer-reviewed scientific literature to synthesize the role of AI/ML in mechanical engineering. The review focuses on publications primarily from the last decade, reflecting the acceleration caused by DL technologies. The synthesis is structured around five critical domains, each representing a distinct phase in the mechanical engineering lifecycle: (A) Design and Generative Engineering, (B) Smart Manufacturing and Process Optimization, (C) Prognostic Health Management and Predictive Maintenance, (D) Control Systems and Dynamic Systems, and (E) Computational Mechanics and Advanced Simulation, including Materials Science. The subsequent Discussion section critically analyzes the interdisciplinary synergies and prevailing challenges, such as model interpretability and the need for robust deployment strategies, before outlining future research trajectories.

2. Methods

The literature search was executed using systematic methodologies adhering to academic standards, utilizing primary engineering and computer science databases, including IEEE Xplore, ScienceDirect, Web of Science, and PubMed Central. The search queries combined broad domain terms with specific AI/ML techniques relevant to mechanical engineering applications. Core keywords and phrases included "AI Mechanical Engineering," "Deep Learning Topology Optimization," "PINN fluid dynamics," "PHM machine learning," "Reinforcement Learning Control Systems," and "Materials Informatics." Boolean operators were employed to ensure comprehensive coverage, focusing on the intersection of advanced AI methodologies and critical mechanical system functions.

The review prioritized peer-reviewed journal articles, high-impact scientific reviews, and conference proceedings to ensure the validity and rigor of the synthesized findings. A temporal filter was primarily applied to include literature published within the last five to ten years, specifically to capture the advancements brought about by the emergence of deep learning, which accelerated research output significantly. Sources focusing purely on theoretical computer science without demonstrated engineering application, or those lacking empirical validation or comparative analysis, were excluded. Emphasis was placed on literature that provided explicit data, performance comparisons, or detailed methodological descriptions pertinent to mechanical engineering problems.

The collected literature was organized thematically based on the five defined domains of mechanical engineering. The synthesis process involved a multi-layered analysis. First, the demonstrated potential and technical capabilities of the AI/ML application (e.g., accuracy, speedup, material efficiency) were quantified based on reported metrics. Second, a critical analysis was performed to identify knowledge gaps, such as the discrepancy between laboratory performance and real-world industrial adoption, and the necessity for physical constraints (e.g., in PINNs). The synthesis framework focused on the underlying algorithmic necessity for each application; for example, analyzing why a hybrid CNN-LSTM structure is superior for prognostic tasks [7], [9] or why evolutionary algorithms must be integrated with reinforcement learning for stable control [12]. This rigorous approach ensured the report moved beyond merely listing applications to providing a nuanced understanding of the fundamental scientific and engineering challenges being addressed by AI/ML. For instance, the systematic mapping of research in mechanism synthesis highlights the necessity of using standardized datasets to allow for comparative analysis of different ML algorithms [23], [24].

3. Thematic Review of Literature (TROL)

A. AI in Mechanical Design and Generative Engineering

1. AI Integration in Computer-Aided Design (CAD) and Automation

The traditional computer-aided design (CAD) paradigm relies on iterative manual input and subsequent computational validation (simulation). AI is fundamentally changing this workflow by transforming CAD from a static modeling tool into a dynamic, real-time design assistant [25]. Modern CAD environments, including commercial platforms such as SolidWorks, Autodesk Fusion 360, and Siemens NX, are increasingly integrating ML algorithms to assist in design creation and refinement [20], [25]. These tools enable engineers to input high-level functional and performance parameters, which the AI then utilizes to generate structurally sound configurations [25]. This automation dramatically reduces the development time required for initial concept iterations and minimizes material waste by efficiently distributing material early in the design phase. Beyond generating basic shapes, AI enhances three-dimensional (3D) modeling by automatically detecting potential design flaws, suggesting corrective measures, and streamlining the transition from concept to prototype [25]. AI also facilitates simulation-driven design, where digital prototypes undergo continuous stress analysis and virtual testing under varied conditions to guarantee reliability before any physical production begins [21], [25].

2. Topology Optimization (TO) and Generative Design (GD) using Deep Learning

Generative engineering, particularly the combination of Topology Optimization (TO) and Generative Design (GD), is a major beneficiary of deep learning advancements [26], [20]. This combined framework is especially powerful when utilized for components intended for Additive Manufacturing (AM). DL models, notably Convolutional Neural Networks (CNNs), are highly effective in handling the high-dimensional spatial data intrinsic to TO problems, such as boundary conditions, load distribution, and material density [2].

The power of integrating DL with TO and GD lies in its ability to explore vast, complex design spaces that were previously intractable for human designers [27], [2]. Generative deep learning models process large datasets to identify subtle patterns in design parameters, allowing them to create novel AM structures that exhibit superior material efficiency and structural integrity compared to designs generated through conventional methods [27]. For example, the use of Conditional Generative Adversarial Networks (GANs) has led to non-iterative TO methods that achieve high pixel-wise accuracy of 83.15%. This predictive capability significantly reduces iterative design cycles, generating numerous optimal design candidates in a fraction of the time required by traditional iterative methods.

To address the inherent computational cost of training and validating DL models within the optimization loop, researchers have introduced "theory-driven" mechanisms, such as adaptive sampling. This technique aims to reduce computational overhead by prioritizing training data points (optimization problem sets) where the predicted topology deviates most significantly from established optimality conditions for TO. By linking the data-driven prediction mechanism back to physical optimality criteria, this method ensures the model provides better predictions of optimal structural compliance for unseen loading scenarios, achieving engineering validity while optimizing computational budget. Furthermore, deep learning models can reduce production time in AM from 50 hours to 30 hours and increase waste reduction by 20% compared to non-DL methods.

3. **Inverse Design and Advanced Generative Models**

The utilization of advanced generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), is central to the movement toward inverse design [11]. Inverse design represents a paradigm shift where the engineer inputs the desired performance metrics (e.g., stiffness, thermal conductivity, frequency response), and the AI system outputs the optimal physical geometry necessary to achieve those metrics [11]. This contrasts sharply with the traditional forward design loop where a geometry is defined and then analyzed for performance.

The applications of this inverse design methodology are diverse, classifying research findings across eight key industry sectors, including aerospace, architecture, automotive, consumer goods, industrial equipment, and robotics. For example, in the automotive sector, inverse design is crucial for generating lightweight yet robust structural components that meet stringent crash safety standards while minimizing material usage.

4. **Mechanism Synthesis and Kinematic Analysis**

Mechanism synthesis, the process of designing mechanical linkages to achieve specific motion requirements, is another domain seeing rapid augmentation through ML [23]. Recent literature demonstrates promising results using deep neural networks for the synthesis of planar mechanisms [23]. By learning the complex relationships between kinematic input parameters and geometric output configurations, ML algorithms can efficiently propose viable mechanisms for target path generation [23].

However, the analysis of current research, often conducted following systematic review protocols like PRISMA, reveals several critical gaps that must be addressed for broader adoption [23]. First, most existing models focus heavily on simple target paths and often neglect crucial real-world design requirements, such as desired velocities, transmission angles, and other complex design constraints [23]. Second, research frequently limits learning to a few selected mechanism types. For ML to be effective in real-world applications, the mechanism type itself must be an output of the synthesis process, not a predefined input [23]. Furthermore, the field requires the creation and utilization of standardized, labeled datasets of mechanisms to allow for meaningful comparison of different ML algorithms and to benchmark performance [23], [24]. Successfully addressing these gaps requires a concerted effort to shift the focus from simple kinematic learning to the generation of mechanisms based on comprehensive, real-world motion task requirements, ensuring that the generated mechanisms are robust and practically manufacturable.

B. AI in Smart Manufacturing and Process Optimization

1. **AI for Real-time Quality Control and Defect Detection**

Artificial intelligence has emerged as a transformative technology in high-tech manufacturing, specifically revolutionizing quality control and process optimization [3], [26]. Quality control, traditionally relying on manual checks or fixed statistical process control, is now being managed by AI-powered real-time monitoring systems. These systems continuously assess production processes using technologies like computer vision and sensor fusion to detect deviations from quality standards and trigger immediate corrective action [3].

The ability of AI to examine complex, multi-dimensional datasets rapidly allows for immediate root cause analysis, identifying underlying quality issues much faster than conventional methods [3]. This capability leads directly to enhanced product consistency, reduced operational costs, and maximized productivity by lowering the necessity for manual intervention and rework [3]. Furthermore, adaptive quality control algorithms can learn from historical data to automatically adjust production parameters in real time, maintaining consistent quality even when faced with changing production conditions or material variability [3]. By developing tailored quality control strategies for each product variant, AI ensures compliance with specified standards and customer expectations.

2. **Process Optimization in CNC Machining and Assembly**

AI plays a critical role in streamlining both material removal (CNC machining) and material addition (3D printing/AM) processes. In conventional manufacturing, AI can optimize component layouts and automate repetitive design modifications, directly improving productivity [25]. For CNC machining, ML models can predict tool wear, optimize cutting parameters (e.g., feed rates and spindle speeds) to reduce cycle time and improve surface finish, and leverage sensor data to detect vibration or chatter in real time.

The successful implementation of AI involves bridging the gap between theoretical understanding and practical implementation [3]. By utilizing advanced technologies such as machine learning, computer vision, and robotics, manufacturers can leverage AI to address long-standing challenges in achieving consistent quality and optimizing process parameters, leading to significant reductions in scrap and waste [3].

3. **AI for Additive Manufacturing (AM) Parameter Prediction**

The complexity of Additive Manufacturing (AM), involving numerous interdependent process variables (e.g., laser power, scan speed, powder characteristics), makes achieving consistent, high-quality builds challenging. AI bridges the gap between design conceptualization and physical realization by optimizing AM process parameters [2], [27]. DL models are trained on data derived from previous successful and failed builds, allowing them to predict the optimal combination of parameters necessary to achieve desired material properties (e.g., porosity, strength, microstructure) and minimize build failures.

The enhanced efficiency achieved through the combined generative design/DL framework, previously discussed in the context of design [27], continues into the manufacturing phase. By learning how to distribute material optimally based on complex constraints, DL models facilitate designs that lead to improved material utilization and waste reduction. The predictive power helps ensure that the generated complex topologies are not only structurally optimal in theory but are also feasible for fabrication via 3D printing [2], [27].

4. **Supply Chain Optimization and Production Scheduling**

Beyond the direct production line, AI extends its influence to enterprise-level operations, notably in optimizing the supply chain and production scheduling [25]. AI-driven digital twins and simulations provide a comprehensive, holistic view of the manufacturing flow, allowing manufacturers to forecast demand, manage inventory, and dynamically re-route materials to optimize efficiency.

For high-throughput, highly customized production systems, Deep Reinforcement Learning (DRL) offers a potent solution [21]. Shortening product development cycles and increasing product diversity pose major challenges, requiring production systems to be highly adaptable and robust to process variations. DRL is increasingly applied because, unlike other ML methods, it operates in direct interaction with the environment using recently collected sensor data, enabling real-time responses to system changes [21]. This adaptability means DRL can optimize complex production systems and provide high throughput, proving superior to conventional scheduling methods by reducing implementation efforts and dependency on human experience [21].

C. AI for Prognostic Health Management (PHM) and Predictive Maintenance (PdM)

1. **Fundamentals of Fault Diagnosis and Prognosis (FDP)**

Prognostic Health Management (PHM) is a proactive maintenance strategy based on Condition Monitoring (CM) data, designed to forecast the future operational health of machinery [4]. PHM systems operate across two primary tasks: fault diagnosis (FDP), which identifies the current state and type of fault, and prognosis, which predicts the timeline to future failure, often quantified as the Remaining Useful Life (RUL). Accurate FDP is crucial for optimizing maintenance schedules, minimizing unplanned downtime, and ensuring the safety of industrial assets. Most PHM applications rely heavily on sophisticated ML and DL techniques trained on large amounts of historical data collected from sensors.

The advent of deep learning has been significant because it integrates the feature engineering and PHM modeling parts, allowing end-to-end fault detection, diagnosis, and prognosis to be performed automatically from raw signals, thereby reducing the need for ad-hoc professional knowledge in signal processing. This advancement includes developing generalized definitions and mathematical formulations for FDP problems compared to earlier work.

2. **Deep Learning Architectures for Remaining Useful Life (RUL) Estimation**

The prediction of RUL requires models capable of analyzing temporal sequences of sensor data, capturing the progressive nature of mechanical degradation over time. Early approaches utilizing traditional machine learning often proved less effective than sequential models due to their limited capacity to handle time series data and long-term dependencies [9].

The development of hybrid Deep Learning architectures has achieved superior results in RUL estimation.

Specifically, the combination of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks forms a powerful hybrid model [7], [9]. In this architecture, the CNN component excels at extracting robust, localized spatial features from the raw sensor data at each time step (e.g., identifying specific frequency components indicative of damage). Following feature extraction, the LSTM component, which is inherently designed for temporal sequence analysis, processes these extracted features over time to predict the degradation trajectory and ultimately, the RUL [7].

This CNN-LSTM fusion model is superior to single LSTM models in predictive performance and consistently outperforms other machine learning algorithms, as demonstrated by its high accuracy on benchmark datasets like the NASA CMAPSS dataset (jet engine sensor data) [7], [9]. For instance, on lithium-ion battery RUL prediction, the CNN-LSTM fusion model achieved a Mean Absolute Percentage Error (MAPE) of 0.36% and a Mean Square Error (MSE) of 0.38e-4, outperforming single LSTM models [8]. The success of this hybridization confirms that mechanical degradation is best modeled by capturing both the localized characteristics of damage (spatial features extracted by CNNs) and the long-term historical evolution of those features (temporal sequencing handled by LSTMs) [7].

3. **Addressing Data Scarcity and Emerging Architectures**

A significant hurdle for PHM adoption is the data dependency of DL models. Many industries struggle to obtain sufficient historical data, particularly data relating to rare or catastrophic failures, making off-line batch analysis difficult [15]. To overcome these challenges, emerging architectures are being investigated:

- **Generative Adversarial Networks (GANs):** GANs are gaining attention in intelligent FDP. They are primarily used for data augmentation, creating synthetic yet realistic fault data that helps train diagnostic models in low-data regimes. This approach improves model robustness and generalization capability, especially for rare fault conditions where empirical data is scarce.
- **Transformers and Graph Neural Networks (GNNs):** These emerging DL architectures are attracting research focus in FDP. GNNs are particularly well-suited for modeling complex relational data, such as interconnected machinery networks, component dependencies, or supply chain graphs. Transformers, known for handling very long sequences effectively, hold potential for complex prognostics over extended operational lifecycles.

4. **Edge-Cloud Architectures for Condition Monitoring**

The operationalization of PHM demands a data architecture capable of handling the volume and velocity of sensor data generated by industrial machinery. The necessary solution involves an integrated edge-cloud infrastructure [15].

At the edge (physically close to the machinery), streaming analysis is performed in real-time. This includes immediate anomaly detection and identification of novel behaviors, allowing rapid awareness of the machinery's health status [15]. Crucially, this edge processing filters and reduces the quantity of data that needs to be permanently stored, mitigating network bandwidth and storage constraints [15].

The cloud component aggregates the filtered data collected from multiple machines across different locations. This high volume of aggregated data is then used to train more accurate diagnostic and prognostic models, benefiting from a statistically significant sample size. The results of these highly accurate models are subsequently deployed back to the edge devices, allowing them to predict health status in real-time [15]. This separation of concerns—real-time monitoring (edge) versus long-term training (cloud)—provides a robust and scalable solution, balancing the need for instantaneous safety monitoring with long-term predictive accuracy in complex, distributed machinery networks. The challenges of this architecture include computational burden, storage, and the need for transfer learning to address evolving environments and equipment fleets.

Table 2: Summary of cloud architectures for condition monitoring and key applications

Architecture	Strength in PHM Context	Weakness/Challenge	Key Application
Convolutional Neural Networks (CNNs)	Excellent for spatial feature extraction from transformed signal data; robust fault diagnosis.	Requires signal preprocessing (e.g., spectrograms); limited inherent temporal memory.	Fault diagnosis from vibration data .
Long Short-Term Memory (LSTM)	Superior ability to capture long-term dependencies in time-series sensor data.	High computational cost for very long sequences; difficulty extracting local features.	Remaining Useful Life (RUL) estimation.9
Hybrid (e.g., CNN-LSTM)	Combines CNN for feature robustness and LSTM for sequence modeling, yielding superior prediction accuracy 7,.9	Increased model complexity and number of hyperparameters.	High-accuracy RUL estimation under variable operating conditions 7,.8
Generative Adversarial Networks (GANs)	Augments scarce, complex fault data, improving robustness and enabling cross-domain transfer .	Challenging training stability and potential for mode collapse.	Synthetic data generation for rare fault conditions .

D. AI in Control Systems and Dynamic Systems

*1. **System Identification (SID) using Neural Networks (NNs)***

Control engineering relies fundamentally on accurate System Identification (SID)—the mathematical modeling of system dynamics based on observed input and output data [5]. The development of AI greatly benefits SID, control, and optimization methods [5]. Traditional SID methods often struggle with complex, highly nonlinear systems or those with significant unknown dynamics. NNs offer a solution by acting as powerful function approximators capable of learning these intricate relationships directly from data, providing a dynamic model necessary for high-performance control design [5].

2. Advanced Control Strategies: Adaptive and Backstepping Control with NN Integration

Neural network integration has substantially enhanced conventional control strategies by adding adaptability and robustness against uncertainty [5]. Traditional adaptive control systems are typically vulnerable to unknown dynamics and uncertainties, making them challenging to apply effectively to complex nonlinear systems. By integrating NNs, controllers become self-tuning, allowing them to estimate and compensate for unknown system parameters or dynamics in real time [5].

Similarly, NNs are critical for improving backstepping control. Backstepping is a recursive design method for stabilizing nonlinear systems, but it often requires finding a suitable control Lyapunov function (CLF), which is extremely challenging for general nonlinear control systems [5]. NNs can be used to approximate the unknown dynamics or components of the CLF, effectively simplifying the design process and expanding the applicability of backstepping to a much wider range of mechanical systems, such as advanced robotics or unstable lightweight structures [5]. The integration of NNs transforms classical control theory, making controllers robust to system uncertainties, which is crucial for modern, high-performance mechanical applications [5].

3. Reinforcement Learning (RL) for Optimal Control and Robotics

Reinforcement Learning (RL) is uniquely suited for designing optimal controllers, as it allows an agent to learn the best sequence of actions to optimize a dynamical system based on performance metrics [5]. RL methodologies are typically categorized based on their dependency on a system model:

- **Model-Based RL:** This approach utilizes a known, or learned, mathematical model of the system dynamics to accelerate the learning process [5]. The model information can be fully used in the RL algorithm to improve sample efficiency [5].

- **Model-Free RL:** In practical applications, obtaining an accurate mathematical model is often difficult. Model-free RL, therefore, directly learns the optimal policy or value function from interacting with the environment, without relying on an explicit system model [5].

Deep Reinforcement Learning (DRL) applies these principles using deep neural networks to approximate the policy or value functions, enabling control over high-dimensional and complex state spaces. DRL is successfully employed in production control, where it demonstrates superior performance over conventional methods in providing high adaptability and robustness to process variations [21]. Furthermore, specific DRL architectures, such as Deep Q-Networks (DQN), are instrumental in addressing fundamental robotic problems like path planning in unknown environments, leveraging self-learning capabilities [22].

Optimal control, often utilizing Approximate Dynamic Programming (ADP), is an important branch of control theory aimed at minimizing a cost function to optimize a system [5].

4. Hybrid Control: AI-Enhanced Model Predictive Control (MPC)

Model Predictive Control (MPC) is a multivariable control algorithm defined by three core components: an internal dynamic model of the process, a cost function defined over a receding time horizon, and an optimization algorithm that minimizes the cost function using control inputs [5]. The internal dynamic model is traditionally based on linear or linearized system approximations, which can limit performance when dealing with highly nonlinear mechanical systems. AI significantly enhances MPC by augmenting or replacing this internal model with sophisticated neural network approximations. For instance, combining Adaptive Neuro-Fuzzy Inference Systems (ANFIS) or various Radial Basis Neural Networks (RBNN) with MPC enhances the predictive capabilities of the internal model [5]. This augmentation allows the controller to accurately forecast system behavior under nonlinear conditions, leading to optimized control in complex applications, such as Eddy Current Dynamometers [5].

Researchers have implemented and validated AI-based MPC on large-sized Eddy Current Dynamometers, comparing techniques like ANFIS, RBNN, SHLNN, and GRNN, and highlighting the novelty of this application in the literature [5]. The integration of AI elevates MPC from relying on approximate models to using highly accurate, learned dynamic models, dramatically improving the controller's ability to handle complex dynamics, thereby enhancing safety and operational efficiency [5].

5. Robustness and Stability in DRL: Evolutionary Reinforcement Learning (ERL)

Despite the profound success of DRL in simulations, its transition to real-world, safety-critical mechanical systems is hindered by inherent limitations: DRL often suffers from brittle convergence properties sensitive to hyperparameters, difficulty with sparse reward environments (temporal credit assignment), and a lack of effective exploration [12]. These challenges severely limit the applicability of pure DRL approaches to critical real-world problems [12].

Evolutionary Algorithms (EAs), black-box optimization techniques inspired by natural evolution, are well-suited to address these weaknesses. EAs provide effective exploration through diverse sets of policies and inherent stability via their population-based approach [12]. Evolutionary Reinforcement Learning (ERL) is a hybrid algorithm designed to leverage the strengths of both paradigms [12]. ERL utilizes an EA population to generate diversified data, which trains the RL agent, and periodically reinserts the RL agent back into the EA population to inject gradient information [12]. This fusion allows ERL to inherit the EA's robustness and exploration capability while leveraging the gradient-based learning efficiency of DRL [12].

ERL represents a necessary technological evolution for adopting DRL in industrial and autonomous settings, overcoming the fragility that often limits pure DRL deployment in critical, real-world control tasks [13].

E. AI in Computational Mechanics and Advanced Simulation

1. Physics-Informed Neural Networks (PINNs): Methodology and Governing Equations

Computational mechanics relies on solving complex Partial Differential Equations (PDEs) that govern physical phenomena (e.g., fluid dynamics, heat transfer, elasticity). This task is traditionally computationally expensive, requiring extensive meshing and time-stepping inherent to methods like FEA and CFD. Physics-Informed Neural Networks (PINNs) offer a revolutionary computational approach that embeds the underlying physical laws directly into the neural network architecture [6].

A PINN typically consists of a deep neural network whose parameters are optimized by minimizing a loss function comprising multiple components: one component minimizes the residual of the governing PDE (ensuring compliance with physical laws within the domain), and the others ensure the network satisfies the boundary conditions (BCs) and initial conditions (ICs) [6], [10]. This structure enables PINNs to solve PDEs without requiring labeled data (unsupervised setting) [10]. Advanced variants, such as those using Deep Backward Stochastic Differential Equations (Deep BSDE) methods, utilize NNs to approximate solutions of high-dimensional PDEs. Integrating PINNs into this framework enhances its capability by explicitly enforcing adherence to governing stochastic differential equations, resulting in more accurate and reliable solutions [10].

2. Applications in Structural Analysis and Elasticity Problems

PINNs have rapidly found applications across various domains of structural mechanics. They have been effectively used for solving problems in linear elasticity and can be extended to highly nonlinear elasticity problems where constitutive equations are complex [6], [10]. Specific applications include solving Kirchhoff plate bending problems under transverse distributed loads and modeling contact mechanics with elastic Winkler's foundations [10]. This methodology demonstrates its capacity to handle complex, domain-specific physical formulations by encoding them mathematically into the optimization landscape of the neural network [6].

The effectiveness of advanced PINN structures, such as the Physics-Informed Point Network (PIPN), has been demonstrated for complex mechanical systems, including incompressible flow, heat transfer, and linear elasticity [10]. Crucially, the ability of deep neural networks (DNNs) to act as physics-informed models allows for the solving of PDEs in unsupervised settings where traditional methods rely on labeled training data [10].

3. Multiscale Modeling: Physics-Constrained DL for Damage and Fracture

Multiscale simulations, which often leverage homogenization theory (e.g., the FE^2 method) to couple macro-scale component behavior with micro-scale material responses, are essential for modeling heterogeneous materials, particularly when analyzing damage and fracture in large components [6]. However, these simulations are prohibitively expensive and memory-intensive [6]. Physics-constrained deep learning (PCDL) models offer a solution by acting as accurate, computationally efficient surrogates for these microscale simulations [6]. A proposed framework utilizes a physics-constrained Recurrent Neural Network (RNN) designed to predict homogenized path-dependent microstructural behaviors [10]. A key requirement for engineering validity is ensuring that the DL model adheres to fundamental physical laws throughout the process. This is achieved by introducing hard constraints:

1. **Thermodynamic Consistency:** A penalty term is added to the loss function based on energy analysis, promoting solutions that do not violate fundamental laws of thermodynamics [10].

2. **Irreversible Damage:** The model incorporates a hard constraint by directly manipulating the temporal variation of the RNN's outputs, strictly accounting for the irreversible nature of damage accumulation processes [10].

By explicitly integrating these hard constraints, the PCDL model ensures that the speedup—demonstrated to be about four orders of magnitude faster than the classic FE^2 approach in terms of CPU hours for online computations—does not compromise the engineering validity or accuracy in capturing complex, irreversible phenomena like elasto-plastic hardening and softening deformations [10].

4. Addressing Spectral Bias and Irregular Geometries (PGCANs)

A known limitation in early DNNs used as PDE solvers is "spectral bias," where the network tends to learn only the low-frequency, smooth characteristics of the solution, missing higher-frequency details crucial for accuracy [10]. Furthermore, applying PINNs to domains with irregular geometries (a common occurrence in mechanical components) often complicates implementation [10]. To address these issues, Parametric Grid Convolutional Attention Networks (PGCANs) have been introduced [6]. PGCANs map the input space to a structured high-dimensional feature space using a parametric grid convolutional encoder [6]. This architecture significantly improves information propagation, especially from the boundaries, and qualitatively and quantitatively assesses spectral bias, demonstrating that PGCANs effectively mitigate this accuracy limitation [10]. Moreover, PGCANs naturally extend to 3D and can handle irregular domain geometries, such as solving the Poisson equation within a torus, substantially increasing the utility and reliability of PINNs for complex mechanical design problems [10]. Importantly, PINN models are highly significant in their capacity to solve inverse problems where conventional methods fail to solve PDEs with unknown parameters [10].

5. AI for Turbulence Modeling and Fluid Dynamics

The accurate modeling of turbulent flows remains one of the greatest unresolved challenges in classical computational mechanics [5]. AI is providing new pathways for modeling turbulence and optimizing thermal transport [5]. Machine learning techniques, including CNNs, RNNs, Deep Reinforcement Learning (DRL), and PINNs, are increasingly applied to turbulence modeling to overcome the limitations of conventional Reynolds-Averaged Navier–Stokes (RANS) equations [5]. While this data-driven approach requires substantial high-quality data for training, ongoing research is refining collection, processing, and generalization methodologies [5].

Specific studies have applied PINNs to solve RANS equations for turbulent flows, such as backward-facing step flow, incorporating turbulence models like the standard k-ε model, and comparing results favorably to Direct Numerical Simulation (DNS) data when sufficient labeled training data (e.g., three to five vertical lines of data) is provided [5]. Furthermore, research on an Adverse Pressure Gradient (APG) boundary layer showed that PINNs effectively model wall pressure and Wall Shear Stress (WSS), with the training cost of the network remaining consistent across different Reynolds numbers [5].

A notable breakthrough demonstrates the capability of DRL to dynamically adjust thermal boundaries in turbulence simulations, achieving a heat transfer enhancement of 38.5%—over 50% better than traditional approaches [5]. Crucially, the AI-derived strategy was distilled into a simple formula that retained effectiveness even in extreme, unencountered conditions, offering a powerful framework for advanced turbulent flow control and optimization in real-world applications [5].

Similarly, ML models are widely utilized in heat exchanger modeling for performance calculation, design optimization, and transient performance predictions, often built upon extensive experimental or numerical datasets [5]. ML is particularly attractive for these multi-physics systems because it provides the computational speed and robustness necessary for optimization studies and real-time control analysis, where rapid iteration and reliable performance estimation are necessary [5]. Specific applications include predicting fouling rates and transient performance under varying flow regimes, see table 3 [5].

Table 3: PINN Methodological Spectrum in Computational Mechanics

PINN/PCDL Variant	Application Domain	Key Technical Challenge Addressed	Core Constraint/Benefit
Vanilla PINNs (vPINNs)	Simple PDEs (e.g., 1D Burgers' equation)	Solving forward/inverse PDEs without requiring labeled data.10	Loss function minimizes PDE residual and boundary conditions.
Physics-Constrained RNNs	Multiscale Damage Modeling (FE^2 surrogate)	Prohibitive computational cost of multiscale simulations.10	Hard constraints for thermodynamic consistency and irreversible damage; four orders of magnitude faster online computation.10
Parametric Grid CNN (PGCAN)	PDEs on irregular domains	Spectral bias and decreased accuracy with increasing PDE complexity.10	Parameterized grid encoding improves information propagation and mitigates spectral bias.10
Deep BSDE Methods	High-dimensional Stochastic Differential Equations	Computational burden of high-dimensionality.10	Embeds physical laws to ensure solutions adhere to governing stochastic equations.10

F. AI in Materials Science and Heat Transfer Engineering

1. Materials Informatics and High-Throughput Property Prediction

The discovery and optimization of new materials with specific mechanical, thermal, and chemical properties are essential tasks in mechanical engineering, traditionally limited by the computational expense of accurate theoretical methods such as Density Functional Theory (DFT). Machine Learning and AI methods have been hailed as the next scientific paradigm, accelerating materials discovery and optimization under initiatives like the Materials Genome Initiative (MGI) [18,19].

ML enables materials informatics to rapidly predict unprecedented thermal properties and other critical characteristics. By learning implicit chemical and geometric knowledge, data-driven modeling approaches provide an efficient means for material property prediction, effectively surrogating computationally demanding high-fidelity simulations [19]. This capability allows researchers to filter through vast hypothetical material spaces quickly, reserving expensive computational resources like DFT for only the most promising candidates, thereby drastically shortening the research and development pipeline.

The use of Graph Neural Networks (GNNs) is rapidly growing in this domain. GNNs are highly relevant as they work directly on the graph representation of molecules and materials, providing full access to structural information. Unlike traditional approaches, GNNs automatically extract node relationships and topology structure information, reducing the cost of manually designing features and eliminating human influence. GNNs have proven effective in predicting material properties, accelerating simulations, and designing new structures.

2. Inverse Materials Design and Crystal Structure Prediction

The paradigm of inverse design extends fully into materials science, moving quickly toward an AI-driven approach. Inverse materials design aims to determine the chemical composition and crystalline structure required to achieve a predefined set of performance metrics.

For crystal structure prediction (CSP), a critical challenge involves efficiently sampling the vast configuration space of atoms, a limitation acutely felt when using conventional DFT-based methods. Methods like DeltaCrystal utilize deep learning to address this limitation. This approach employs a deep residual neural network to learn and predict the atomic distance matrix for a given material composition, based on patterns found in known structures. Subsequently, this predicted matrix is used to reconstruct the 3D crystal structure via optimization algorithms, such as genetic algorithms. Through this approach, the model effectively learns implicit interatomic relationships, demonstrating superior structure prediction performance for more complex crystals compared to global optimization-based CSP methods. This capability is instrumental in accelerating the discovery of novel alloys and composites needed for advanced mechanical systems operating under extreme conditions.

3. ML for Heat Exchanger Modeling and Thermal System Optimization

Heat exchangers are integral components in countless mechanical systems, from power plants to air-conditioning units, and their modeling for performance calculation, design optimization, and control is crucial. Traditionally, their analysis relied on theoretical, analytical, or numerical methods, which often lack the computational speed and robustness required for modern real-time optimization studies.

Machine learning models, trained on experimental or high-fidelity numerical data, significantly improve the state-of-the-art simulation approaches by offering high prediction capability as a regression tool. ML is used to predict crucial factors such as fouling rates, thermodynamic properties of non-ideal fluids, and transient performance under varying flow regimes [19,20]. The speed and robustness offered by ML models are essential for optimization studies and control analysis, where rapid iteration and reliable performance estimation are necessary. However, the successful deployment of these models relies heavily on the quality and integrity of the database used for training, as well as the appropriate selection and implementation of the ML algorithm.

4. Discussion

A. Critical Synthesis of Findings and Interdisciplinary Synergies

The literature overwhelmingly demonstrates that AI/ML is not merely optimizing isolated processes within mechanical engineering, but is actively facilitating the convergence and creation of powerful new interdisciplinary loops across previously distinct domains.

1. Design-Simulation Integration

A primary synergy exists between computational simulation and generative design [27]. Traditional generative design relies on iterative validation using expensive computational methods. The development of Physics-Informed Neural Networks (PINNs) and physics-constrained deep learning surrogates (Section III.E) fundamentally changes this dynamic. By creating surrogate models that are four orders of magnitude faster than conventional multiscale methods while maintaining thermodynamic consistency [10], PINNs allow designers to integrate physically reliable, high-fidelity simulation checks directly into the rapid, iterative generative loop. This integration enables designers to iterate on designs with complex material behaviors, such as fracture and damage, in near real-time, thereby ensuring that computationally generated components are structurally sound and physically realizable. The rapid growth of generative AI, including models like Diffusion Models and Large Language Models (LLMs), suggests this design-simulation loop will only accelerate [20].

2. Control-PHM Synergy

A vital link is forged between Prognostic Health Management (PHM) and advanced control systems. Highly accurate prognostic models, particularly those based on hybrid CNN-LSTM architectures [7], provide precise estimations of future degradation and Remaining Useful Life (RUL). This predictive output can be directly fed into AI-enhanced Model Predictive Control (MPC) systems. Instead of relying on a static, nominal model, the AI-enhanced MPC can adjust operational parameters (e.g., rotational speed, load limits) based on the predicted future state of degradation. This synergy allows for truly optimized and resilient system operation, enabling systems to run safely up to the exact point of predicted failure, maximizing asset utilization, and integrating maintenance tasks seamlessly with operational planning.

3. Material-Design Loop Closure

AI closes the loop between materials science and product design, rapidly accelerating the material development cycle. Materials informatics models (Section III.F), such as DeltaCrystal and GNNs, quickly identify and predict novel material properties and crystal structures. This inverse materials knowledge [11] immediately informs the Generative Design and Topology Optimization process (Section III.A). Instead of designing a component and then selecting an existing material, the process becomes holistic: the system can co-design the component geometry and the necessary material characteristics simultaneously, leading to components optimized for novel, high-performance alloys and composites that were previously inaccessible through conventional, linear R&D pipelines.

B. Key Challenges and Limitations in AI Integration

1. Data Dependencies: Quality, Quantity, and Standardization

Deep learning's efficacy is predicated on the availability of large, high-quality, labeled datasets. In mechanical engineering, this dependency presents multiple hurdles. First, real-world data, particularly sensor data from industrial machinery, is often fragmented, noisy, and unlabeled. Second, achieving robust Prognostic Health Management (PHM) requires data on rare or catastrophic failures, which, by definition, are difficult and costly to obtain in sufficient quantity for training robust DL models [15].

Furthermore, for specialized domains like mechanism synthesis, the lack of standardized, labeled datasets impedes comparative research and broad application [23,24]. Researchers must currently spend significant effort creating proprietary datasets, limiting the ability to compare the efficiency and generalization capacity of different ML algorithms across institutions. The transition to widespread ML adoption demands industry-wide protocols for data collection, labeling, and sharing, potentially mitigated through synthetic data generation techniques like GANs.

2. The Interpretability and Explainability Crisis (XAI)

Perhaps the most critical barrier to deploying AI in safety-critical mechanical systems is the "black box" problem [14]. The scale and inherent nonlinearity of modern AI models, particularly large deep learning architectures used in generative design [20] or autonomous control, often exceed human comprehension [14]. This creates a widening gap between machine knowledge and human understanding of the system's decision-making process.

For mechanical engineering, where failure can result in catastrophic outcomes (e.g., aerospace components, high-speed robotics), Explainable AI (XAI) is paramount [16]. Generative AI and Large Language Models (LLMs) pose significant interpretability challenges due to their complex, large-scale operations and technical barriers like feature superposition. Without the ability to interpret a design or control decision, engineering trust is undermined. The inability to articulate why an autonomous system made a critical decision poses severe operational and regulatory barriers [14]. XAI is essential not just for building trust, but for compliance with regulatory imperatives, establishing legal liability, and performing effective root cause analysis after a failure occurs [14]. As AI evolves toward autonomous systems, interpretability becomes crucial for understanding and directing their decision-making processes.

3. Robustness, Regulatory, and Ethical Concerns

While Deep Reinforcement Learning (DRL) demonstrates superior performance in controlled environments, there remains a critical challenge in transferring these findings to real-world production systems [21]. DRL often exhibits fragile convergence properties and unpredictable behavior when faced with unforeseen conditions or environmental shifts [12]. This lack of demonstrated reliability and the uncertainty regarding safety aspects hinder the practical implementation of DRL in high-throughput manufacturing and dynamic control, necessitating hybrid approaches like ERL for increased stability [12,13]. Furthermore, the general adoption of automation, including ML, remains inconsistent across academic evidence synthesis, suggesting fundamental barriers related to guidance, user awareness, and trust in reliability must be addressed for broader engineering adoption.

More broadly, the review emphasizes the imperative for responsible innovation. The adoption of AI must align with broader societal aspirations for progress, sustainability, and inclusivity [16]. Establishing clear regulatory frameworks is necessary to govern AI agents operating in autonomous mechanical systems, ensuring that ethical considerations are addressed and that AI integration does not lead to unintended consequences, particularly those impacting safety and reliability.

4. Computational Cost and System Complexity

Although methods like PINNs and physics-constrained deep learning models offer remarkable speed-ups during online computation (deployment) [10], the initial cost associated with data generation, model training, and the specialized hardware (GPUs) required remains substantial. This challenge is also noted as a major barrier in the growth of generative design [20]. Moreover, the systems themselves are becoming increasingly complex. The superior performance achieved through hybrid architectures, such as the CNN-LSTM for RUL estimation 7 or the necessity of Edge-Cloud infrastructure for PHM deployment [15], introduces significant systems integration complexity (see table 4). Deploying and maintaining these heterogeneous architectures requires specialized expertise across multiple domains, including network engineering, cloud infrastructure management, and advanced DL knowledge, presenting a barrier to entry for smaller enterprises.

Table 4: Summary of Critical Challenges in AI/ML Deployment in Mechanical Engineering

Challenge Domain	**Core Issue**	**Impact on ME Systems**	**Supporting Source**
Data and Reliability	Lack of standardized, labeled datasets; reliance on rare failure data.	Limits generalizability and robustness of models, impeding industrial adoption and creating operational risk 15,.23	Need for standardized mechanism synthesis data 23; data scarcity in industrial PHM.15
Interpretability (XAI)	Scale and nonlinearity of deep models create non-transparent decisions ("black box").	Hinders regulatory compliance, engineering trust, and effective failure investigation 14, .	Complexity defies traditional interpretability; impacts safety and legal liability 14, .
Transition to Real-World	Difficulty transferring high-performance lab results (especially	Safety risks and unknown reliability under unforeseen operating conditions 21, .	Focus on safety and reliability needed for production implementation 21; ERL for stability 12,.13

	RL) to physical systems.		
Physical Consistency	Purely data-driven models risk violating fundamental physical laws.	Solutions may be non-physical or numerically unstable in critical simulation/design 6,.10	Physics-informed constraints are essential for multiscale accuracy and validity.10

C. Future Perspectives and Research Trajectories

The evolution of AI in mechanical engineering points toward several definitive future research trajectories focused on robustness, integration, and human-machine collaboration.

1. Hybridizing Mechanistic and Data-Driven Models

The future standard for high-stakes mechanical simulation and design will be defined by the seamless hybridization of mechanistic domain knowledge and data-driven learning. The current success of Physics-Informed Neural Networks (PINNs) and physics-constrained deep learning (PCDL) confirms this trend [6],[10]. Future research must focus on moving beyond external loss function penalties to fully integrating domain knowledge—such as thermodynamic principles or constitutive models—directly into the structure and initialization of the neural network architecture. This mechanistic integration will guarantee solutions that are inherently physically consistent and robust against extrapolation errors, making the AI tools viable for critical certification processes.

2. Development of Next-Generation Architectures

The utility of emerging network architectures, particularly Graph Neural Networks (GNNs) and Transformer models, is expected to grow significantly. GNNs are uniquely positioned to analyze complex relational data, modeling intricate component dependencies in large mechanical systems, optimizing supply chain networks, or providing detailed fault diagnosis across interconnected machinery [18],[19]. Furthermore, Generative AI, beyond its immediate use in design [20], will likely expand into new areas, such as synthesizing rapid hypotheses, summarizing complex research findings, and providing automated knowledge extraction from vast bodies of unstructured engineering documentation, potentially leveraging models like diffusion models and LLMs [20].

3. Autonomous Systems and Evolutionary Reinforcement Learning (ERL)

The trajectory toward truly autonomous design, manufacturing, and control systems will be enabled by robust RL variants like Evolutionary Reinforcement Learning (ERL). ERL's ability to maintain stability and explore complex solution spaces overcomes the major fragility issues associated with pure DRL, paving the way for adaptive production lines and sophisticated robotic control [12]. Critical future research must prioritize the development of Verified and Validated (V&V) methodologies specifically tailored for these autonomous, learning systems. These V&V processes must ensure that safety margins are preserved and guaranteed, even as the system continuously learns and adapts its policy in real-time operating conditions. The potential of DRL to achieve optimal real-world performance, such as realizing over 50% better heat transfer enhancement in turbulent flow control [21], mandates continued research into its reliability and safe deployment.

4. Promoting an AI-Ready Workforce

The successful integration of Scientific AI (SciAI) into mechanical engineering demands a profound shift in workforce capability and structure [17]. It necessitates fostering an AI-ready workforce capable of understanding both fundamental mechanical principles and sophisticated computational intelligence techniques [17]. The field requires deep interdisciplinary collaboration, breaking down the historical silos between computer science, electrical engineering, and traditional mechanical engineering departments. Only through collaborative, integrated expertise can the full potential of AI/ML be responsibly harnessed for progress and sustainability.

5. Conclusion

Artificial Intelligence and Machine Learning have irrevocably transformed the landscape of mechanical engineering, transitioning the discipline from an era dominated by analytical and high-cost computational analysis to one defined by data-driven speed, optimization, and generative intelligence. The evidence synthesized across design, manufacturing, PHM, control, and computational mechanics confirms that AI provides tools for achieving levels of efficiency and complexity previously unattainable, whether through the four orders of magnitude speed-up offered by physics-constrained surrogate models 10, the superior accuracy of hybrid CNN-LSTM networks in RUL prediction, or the ability of generative models to design components optimized for novel materials. However, the analysis underscores that the next decade of progress hinges not solely on algorithmic power, but critically on addressing systemic challenges. The imperative is to move beyond proof-of-concept demonstrations toward robust, reliable, and deployable systems. This requires solving the data scarcity problem through architectural solutions like Edge-Cloud integration, mitigating the inherent fragility of autonomous control through hybrid ERL methods, and, most importantly, establishing rigorous Explainable AI (XAI) frameworks to satisfy regulatory demands and build engineering trust in safety-critical applications. In conclusion, the full potential of AI and ML within mechanical engineering can only be realized through responsible innovation, stringent ethical considerations, and robust interdisciplinary collaborations. By prioritizing physical consistency, trustworthiness, and systematic reliability, the engineering community can ensure that this transformative technology aligns with societal aspirations for progress and sustainability, establishing AI as the central pillar of advanced mechanical systems design and operation.

References

1. Lamba, "Artificial Intelligence and Machine Learning in Mechanical Engineering," *J. Adv. Res. Mech. Eng. Tech.*, vol. 10, no. 3&4, pp. 28–34, 2023. [Online]. Available: https://journals.indexcopernicus.com/publication/4020960
2. Willson, "The Impact of AI in Mechanical Engineering," *Fictiv*, Mar. 20, 2025. [Online]. Available: https://www.fictiv.com/articles/ai-in-mechanical-engineering
3. Agrawal, Choudhary, and Wolverton, "Scientific AI in Materials Science: A Path to a Sustainable and Scalable Paradigm," *Proc. Natl. Acad. Sci. USA*, vol. 118, no. 10, 2021. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7919383/
4. Parmar, "Artificial Intelligence in High-tech Manufacturing: A Review of Applications in Quality Control and Process Optimization," *ResearchGate*, 2025. [Online]. Available: https://www.researchgate.net/publication/388194318
5. Al-Dulaimi, Zabihi, Asif, and Mohammadi, "Hybrid Deep Neural Network Model for Remaining Useful Life Estimation," *Concordia University*, 2025. [Online]. Available: https://www.researchgate.net/publication/387321873
6. Khan, Khan, and Khan, "Predictive Maintenance: A Novel Framework for a Data-Driven, Semi-Supervised, and Partially Online Prognostic Health Management Application in Industries," *Appl. Sci.*, vol. 11, no. 8, p. 3380, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/8/3380
7. Panzer and Bender, "Deep Reinforcement Learning in Production Systems: A Systematic Literature Review," *Int. J. Prod. Res.*, vol. 60, no. 13, pp. 4316–4341, 2022. [Online]. Available: https://doi.org/10.1080/00207543.2021.1973138
8. Wikipedia contributors, "Physics-informed neural networks," *Wikipedia*, Oct. 3, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Physics-informed_neural_networks
9. Hosseinmardi, "Physics-informed Deep Models in Computational Mechanics," Ph.D. dissertation, Univ. of California, Irvine, 2024. [Online]. Available: https://escholarship.org/content/qt3507s96d/qt3507s96d.pdf
10. Khadka and Tumer, "Evolution-Guided Policy Gradient in Reinforcement Learning," in *Proc. NeurIPS*, 2018. [Online]. Available: https://papers.neurips.cc/paper/7395-evolution-guided-policy-gradient-in-reinforcement-learning.pdf
11. Kumar, Singh, and Sharma, "Integrating Deep Learning with Generative Design and Topology Optimization for Efficient Additive Manufacturing," *ResearchGate*, 2025. [Online]. Available:

https://www.researchgate.net/publication/385709780

12. Zhang, Liu, and Wang, "Topology Optimization via Machine Learning and Deep Learning: A Review," *J. Comput. Des. Eng.*, vol. 10, no. 4, pp. 1736–1757, 2023. [Online]. Available: https://academic.oup.com/jcde/article/10/4/1736/7223974
13. Suresh, Patel, and Chen, "Topological Optimization and Generative Design for Additive Manufacturing: Software, Trends, and Opportunities," *J. Mech. Des.*, vol. 148, no. 1, p. 014501, 2026. [Online]. Available: https://asmedigitalcollection.asme.org/mechanicaldesign/article/148/1/014501/1218561
14. Khan, Khan, and Ahmed, "Machine Learning Applications for the Synthesis of Planar Mechanisms—A Comprehensive Methodical Literature Review," *ResearchGate*, 2025. [Online]. Available: https://www.researchgate.net/publication/385783622
15. Chen, Zheng, Chen, and Zhao, "Remaining Useful Life Prediction of the Lithium-Ion Battery Based on CNN-LSTM Fusion Model and Grey Relational Analysis," *Electron. Res. Arch.*, vol. 31, no. 2, pp. 633–655, 2023. [Online]. Available: https://www.aimspress.com/article/doi/10.3934/era.2023031
16. Qiu, Cui, Ping, Shan, Li, Bao, and Xu, "Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review," *Sensors*, vol. 23, no. 3, p. 1305, Jan. 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/3/1305
17. Khan, Khan, and Ahmed, "Review on System Identification, Control, and Optimization Based on Machine Learning," *Mathematics*, vol. 13, no. 6, p. 952, Mar. 2025. [Online]. Available: https://www.mdpi.com/2227-7390/13/6/952
18. Zhang, Liu, and Wang, "Path Planning in an Unknown Environment Based on Deep Reinforcement Learning with Prior Knowledge," *ResearchGate*, 2025. [Online]. Available: https://www.researchgate.net/publication/355761412
19. Wikipedia contributors, "Model Predictive Control," *Wikipedia*, Oct. 3, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Model_predictive_control
20. Alshammari, Alqahtani, and Alzahrani, "Model Predictive Control Coupled with Artificial Intelligence for Eddy Current Dynamometers," *Comput. Syst. Sci. Eng.*, vol. 44, no. 1, pp. 1–15, 2025. [Online]. Available: https://www.techscience.com/csse/v44n1/48072/html
21. Khan, Khan, and Ahmed, "Review of Challenges and Opportunities in Turbulence Modeling: A Comparative Analysis of Data-Driven Machine Learning Approaches," *J. Mar. Sci. Eng.*, vol. 11, no. 7, p. 1440, Jul. 2023. [Online]. Available: https://www.mdpi.com/2077-1312/11/7/1440
22. Zhang, Li, and Wang, "Deep Reinforcement Learning Control Unlocks Enhanced Heat Transfer in Turbulent Convection," *Proc. Natl. Acad. Sci. USA*, vol. 121, no. 4, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12452834/
23. Singh, Patel, and Joshi, "Recent Advances in the Applications of Machine Learning Methods for Heat Exchanger Modeling—A Review," *Front. Energy Res.*, vol. 11, p. 1294531, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenrg.2023.1294531/full
24. Roy, "Predicting Material Structures and Properties Using Deep Learning and Machine Learning Algorithms," M.S. thesis, Univ. of South Carolina, 2023. [Online]. Available: https://scholarcommons.sc.edu/etd/7481/
25. Zhao, Li, and Chen, "Artificial Intelligence and Generative Models for Materials Discovery—A Review," *arXiv preprint*, arXiv:2508.03278, Aug. 2025. [Online]. Available: https://arxiv.org/abs/2508.03278
26. Wang, Liu, and Zhang, "Machine Learning Aided Design and Optimization of Thermal Metamaterials," *Proc. Natl. Acad. Sci. USA*, vol. 121, no. 2, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11009967/
27. Deloitte Insights, "The Challenge of AI Interpretability," *Deloitte US*, Oct. 3, 2025. [Online]. Available: https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/articles/ai-interpretability-challenge.html