machines

MDPI

Article

A Digital-Twin-Enabled AI-Driven Adaptive Planning Platform for Sustainable and Reliable Manufacturing

Mingyuan Li [1,2], Chun-Ming Yang [3,*], Wei Lo [4] and Yi-Wei Kao [5]

[1] Institute for Six-Sector Economy, Fudan University, Shanghai 200433, China; 2018210020@gxufe.edu.cn
[2] School of Business Administration, Guangxi University of Finance and Economics, Nanning 530007, China
[3] School of Economics and Management, Dongguan University of Technology, Dongguan 523808, China
[4] College of Economics and Management, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 000p250010@fafu.edu.cn
[5] Department of Applied Statistics and Information Science, Ming Chuan University, Taoyuan City 333321, Taiwan; ywkao@mail.mcu.edu.tw
* Correspondence: 2020812@dgut.edu.cn

Abstract

The manufacturing systems face growing demands due to the instability of the market, the demanding sustainability policies, and the high rate of old equipment, but traditional planning structures are mostly fixed and deterministic, leading to the inefficiency of joint optimization of operational stability and environmental sustainability in unpredictable situations. This research proposed and empirically tested an artificial-intelligence-based adaptive planning platform, which combines a physics-based Digital Twin (DT) and a Pareto-conditioned Multi-Objective Proximal Policy Optimization (MO-PPO) algorithm to be able to co-optimize reliability and sustainability indicators in real-time. The platform reinvents manufacturing planning as a Constrained Multi-Objective Markov Decision Process (CMDP), optimizing an Overall Equipment Effectiveness (OEE) and energy carbon intensity as well as material waste, and strongly adhering to operational restrictions. The study utilizes a four-layer cyber–physical architecture, which includes an edge-based data acquisition layer, a high-fidelity stochastic simulation engine that is calibrated via Bayesian inference, a graph attention network-based state-encoding layer, and a closed-loop execution loop that runs with 60 s long planning cycles. In this study, a statistically significant enhancement was shown in 10,000 stochastic simulation experiments and a 12-week industrial pilot deployment: 96.8% schedule performance, 84.7% OEE, 16.5% cut in specific energy usage (2.38 kWh/kg), 17.1% reduction in material-waste rate (6.8%), and 21.4% enhancement in carbon effectiveness, outperforming all baseline strategies ($p = 0.001$). The analysis showed that there was a surprising synergistic correlation between waste minimization and OEE enhancement ($r = -0.73$), and 34.1% of overall OEE improvement could be explained by sustainability strategies. This study provides a robust framework for adaptive, resilient, and eco-friendly manufacturing processes in line with Industry 5.0 ideologies.

Keywords: adaptive planning; digital twin; multi-objective reinforcement learning; proximal policy optimization; sustainable manufacturing; predictive maintenance; operational reliability

Check for updates

Academic Editors: Qinglin Qi, Lin Wang, Matteo De Marchi and Christophe Danjou

Received: 29 December 2025
Revised: 27 January 2026
Accepted: 2 February 2026
Published: 9 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Manufacturing systems are undergoing radical change due to digitalization, market instability, and the growing need for sustainability. The industry contributes about 37% of

the world's energy use and close to 25% of CO_2 emissions linked to energy, which places manufacturing at the crossroads of decarbonization policies and regulatory interventions. At the same time, heightened competition and reduced product life cycles require that production systems have a high degree of reliability, resource efficiency, and adaptive responsiveness when faced with uncertainty [1–3].

Most production planning and control systems are also still deterministic, even with the advancements in technology. Enterprise Resource Planning (ERP), Material Requirement Planning (MRP), and Advanced Planning and Scheduling (APS) models are based on hard-determined optimization assumptions that do not represent the variability of demand in the real world, stochastic machine failures and unexpected downtime, which usually leads to the rapid degradation of the plan, higher energy consumption, material wastes, and disruptions in delivery [4]. Even though artificial intelligence has enhanced demand forecasting and predictive maintenance, these technologies are normally implemented in disjointed ways and optimized without considering scheduling and sustainability goals, which results in disjointed decision-making and unintended trade-offs [5]. At the same time, regulatory schemes like the EU CBAM and ISO 50001 [6], along with policies of the circular economy, are increasingly imposing combined energy consumption cuts, carbon intensity, and waste of materials on aging industry equipment [7–10]. The current planning solutions are poorly adapted to these needs because of the optimization of all these aspects as being static, the disconnecting of objectives because of siloing, and mismatches between the models and plants because of the dynamics of time-varying energy and degradation. However, a significant research gap remains: the lack of a framework. There is no framework that can be used to real-time, closed-loop co-optimize operational reliability, environmental sustainability, and material efficiency in a high-fidelity cyber–physical environment. The existing methods are unified in terms of individual dimension consideration, simplified models, and not proven to be industrial-scale, whereas Multi-Objective Reinforcement Learning (MORL) methods are challenged by issues of sample inefficiency, policy instability, and limited explainability [11–15]. The existence of this gap is the reason why an integrated, physics-informed adaptive planning platform should be developed that will be able to guide reliable and sustainable manufacturing decision-making in real-time.

This study was motivated by the increasing gap between the theoretical developments in Artificial Intelligence (AI) and the real manufacturing requirements. Although reinforced learning, Digital Twins (DTs), and predictive maintenance each showed promise, their disjointed implementation did not allow benefits to the system. The growing regulatory pressure on carbon emissions and the economic consequences of unplanned downtimes required a flexible planning solution that would be able to learn non-intuitive operating plans that would balance sustainability with reliability [16–18]. The motivation behind this study was the necessity to create a bridge between simulation and reality, to turn sustainability into a co-optimization goal, and provide deployable intelligence in real manufacturing settings.

The modern manufacturing system produces according to fluctuating demand, strict sustainability policies, and out-of-date equipment, which requires constant adjustment over fixed planning. The existing systems of planning and control do not have the ability to optimize sustainability and reliability in real-time together, which leads to the rescheduling of energy-consuming nature, waste, and the loss of reliability when disruptions occur. These inefficiencies have a direct negative effect on regulatory compliance, operational resilience, and long-term competitiveness. Without incorporated adaptive intelligence, the manufacturers must make a tradeoff between reliability and sustainability, thus compromising the two objectives [19]. The present study addresses the lack of a cohesive planning

paradigm that enables data-driven, intelligent, and explainable decision-making across production, maintenance, and sustainability objectives.

This study proposes the inaugural adaptive planning platform that is based on AI and represents manufacturing planning as a Constrained Multi-Objective Markov Decision Process (CMDP) and optimizes it using physics-informed, Pareto-conditioned reinforcement learning within a high-fidelity DT. In contrast to past studies, the proposed solution can optimize OEE, energy carbon intensity, and material waste in real time simultaneously, maintain policy diversity, and is also shown to exhibit stable simulation-to-reality transfer when deployed in industry. The results are one of a kind that sustainability and reliability goals can be complementary to each other, as opposed to being hostile, thus enhancing the knowledge of manufacturing theory and industrial practice [20].

The initial purpose of this study was to design, develop, and test an artificial-intelligence-based adaptive planning platform, which can co-optimize sustainability and reliability in manufacturing operations in an uncertain environment in real-time. The specific objectives were to:

(1) Engineer a single, cohesive cyber–physical design incorporating a high-fidelity DT of physics-informed DT with state-of-the-art RL algorithms in closed-loop manufacturing planning.
(2) Formulate manufacturing planning as a constrained multi-objective decision problem, which optimizes operational reliability measures (e.g., OEE and schedule adherence), at the same time considering the sustainability measures (e.g., energy consumption, carbon intensity, and material waste).
(3) Develop a scalable MORL algorithm that maintains policy diversity, ensures satisfaction of constraints, and can dynamically respond to real-time disruptions.
(4) Evaluate the usefulness and strength of the suggested platform using a thorough stochastic simulation experiment and industrial pilot implementations in diverse disruption conditions.
(5) Analyze the trade-offs and synergies of sustainability and reliability and provide actionable information to the decision-makers to inform both a strategic and operational plan.

The study shows that AI-based adaptive planning can be used to jointly optimize sustainability and reliability in the manufacturing domain in real time. It provides a scalable and practical framework that breaks the constraints of the non-reconfigurable planning and enables energy-efficient, dependable, and resilient manufacturing processes.

2. Literature Review

2.1. Adaptive Manufacturing Planning and AI-Based Decision Systems

The aspect of adaptive manufacturing planning has received growing attention in the literature on Industry 4.0 and Industry 5.0 due to increasing system complexity and uncertainty. According to Sony and Naik [21], successful digital transformation must involve AI-based systems of decisions, which would be responsive in real-time, cross-functional, and human-centric implementation. A systematic survey by Soori et al. [22] reported that AI-based decision support systems achieved improvements of approximately 10–25% in system responsiveness and decision accuracy, while most existing solutions remain functionally isolated and poorly integrated across production, maintenance, and sustainability domains. Concentrating on reinforcement learning (RL), Esteso et al. [23] found that makespan and work-in-process were reduced by 12–18% under stochastic demand conditions. Parallel to scheduling optimization, recent deep learning advances have significantly improved process-level prediction, such as the use of hybrid Dung Beetle Optimization–optimized one-dimensional Convolutional Neural Network with Long

Short-Term Memory (DBO-1DCNN-LSTM) algorithms for surface roughness forecasting in grinding [24] and Convolutional Neural Network (CNN) for predicting geometric profiles in electrochemical machining [25]. While these methods enhance local process fidelity, Modrak et al. [26] found that Deep Reinforcement Learning (DRL) is superior.

The recent research has introduced AI-based planning to decentralized and real-time decision-making. Achamrah and Attajer [27] offered a MORL model of sustainable maintenance choices, but based on fixed preference weights, and it did not consider the real-time interactions with the scheduling. Johnson et al. [28] have reported up to 20% tardiness reduction in robotic assembly cells using multi-agent DRL, and del Real Torres et al. [29] found that there were unresolved issues on training stability, safety assurance, and industrial adoption. Vespoli et al. [30] and Chang et al. [31] made progress towards dynamic scheduling and work-in-process control, which enhanced robustness under disturbances, but omitted energy and degradation dynamism. Handling transparency remains a critical challenge; while recent studies have employed Explainable AI (XAI) methods, like Shapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) to interpret black-box models in complex machining processes [32], Moosavi et al. [33] demonstrated the explainable AI to promote the operator trust at the cost of computational complexity. In addition, an inverse-kinematics-based modeling method was proposed by Lienenlueke et al. [34] to estimate the error in path inaccuracies during machining operations operated by the robot to enhance the geometric accuracy, although it does not consider adaptive planning or real-time optimization, and Del Gallo et al. [35] and Zhang et al. [36] emphasized the coordination overhead and simplified reliability modeling as the continuing drawbacks. Even though Green et al. [37], Li et al. [38], and Kasie et al. [39] report benefits of sustainability and intelligent manufacturing by using AI, current AI-based planning systems are disjointed, lack multi-objective capabilities, and fail to co-optimize reliability and sustainability in real-time.

Recent developments in deep learning have also shown good performance in predicting process-level manufacturing performance, including surface roughness, dimensional deviation, tool wear development, and cavity profile evolution. Convolutional and recurrent neural architectures have been used to successfully process high-frequency sensor signals to learn microscale process behavior to enhance quality prediction and defect prevention. The advancements complete the system-level decision intelligence with the provision of fine-grained process awareness and locate the modern manufacturing intelligence in a multiscale AI paradigm, bridging process-level prediction and shop-floor planning and control. In this setting, the presented DT framework focuses on adaptive decision-making on a system-wide level without contradiction to the process-level learning models in the context of future extension.

2.2. Sustainable Manufacturing and Reliability-Oriented Operations

Ding et al. [40] integrated Remaining Useful Life (RUL) prediction with multi-agent DRL in the joint production and maintenance scheduling with 1824% decrease in unplanned downtime, but energy conditions were held constant, and environmental goals were not explicitly formulated. In a larger context, Scharmer et al. [41] compared sustainable manufacturing frameworks and indicated the increasing participation of environmental and social aspects, but noted that operational decision-making is mostly unchanged. It has been established that predictive maintenance is among the foundational reliability enablers, and Carvalho et al. [42] have reported prognostic accuracies of 85–90% with the help of machine learning. However, industrial deployment often faces challenges related to sensor placement and environmental variability; for instance, recent work in Laser Powder Bed Fusion (LPBF) has demonstrated that accounting for acoustic emission

source motion and sensor positioning via frequency analysis is critical for robust defect detection [43]. Similarly, Selcuk [44] has noted reliability improvement accompanied by enduring challenges with respect to industrial integration and deployment. In the same manner, Beier et al. [45] established that Industry 4.0 technologies increase the sustainability performance, but adoption in maintenance and production functionalities is disproportionate, and Lee et al. [46] established that industrial artificial intelligence enhances reliability management and sustainability is a by-product.

Various studies have been completed to identify frameworks that explicitly relate maintenance, sustainability, and system robustness. Karim [47] analyzed the digital and lean-based maintenance strategies and their contribution to the enhancement of reliability and sustainability, but with a stronger emphasis on offline or non-adaptive strategies. He et al. [48] proposed cost- and reliability-oriented predictive maintenance models of cyber manufacturing systems; however, real-time adaptability was restricted. To discuss sustainability closer to heart, Abadi et al. [49] considered the impact of artificial intelligence and DTs on sustainable production systems, whereas Prasara-A and Gheewala [50] supported the overall sustainability–technology relationship using social sustainability evaluation. Mourtzis et al. [51] highlighted the necessity to look at reliability and resilience jointly, but Leng et al. [52] demonstrated that blockchain-based traceability enhances lifecycle sustainability with indirect operational impacts. The other review by Vrignat et al. [53] and Ghadge et al. [54] also confirmed that Industry 4.0 technologies have a positive impact on sustainability, both in manufacturing and supply chains, whereas the article by Lee et al. [55] showed that smart analytics can improve service innovation and operational reliability, but did not directly address the issue of sustainability–reliability co-optimization. Overall, these works demonstrate how adaptive, real-time frameworks that would coordinate the optimization of reliability and sustainability on the operational level are needed.

Despite the proven effectiveness of learning-based decision systems, their adoption in industrial applications is often limited by poor interpretability and the widespread perception of artificial intelligence as a black box. To mitigate this issue, XAI techniques have been increasingly applied to manufacturing analytics to improve model transparency. Representative methods such as SHAP for feature attribution and Grad-CAM for visual interpretation enable the identification of influential sensor features, enhance operator trust, and support diagnostic reasoning under abnormal operating conditions. Nevertheless, the practical implementation of XAI in industrial environments remains challenging due to sensor noise, spatial correlations, and heterogeneous data sources. These challenges highlight the necessity of robust feature extraction strategies and well-structured data acquisition pipelines within cyber–physical manufacturing systems.

2.3. Digital Twins and Multi-Objective Reinforcement Learning for Cyber–Physical Manufacturing

Making decisions in cyber–physical manufacturing systems using adaptive control is of considerable importance, and the combination of DTs and RL has become a prominent approach to this aim. Zhang et al. [56] proposed a concept of Twins Learning, which merges DTs with RL as real-time shop-floor scheduling and proved to be more responsive to disturbances, but sustainability goals were not addressed. Xia et al. [57] demonstrated that DT environments safely train deep RL agents and can transfer simulation to reality, but their models were not as faithful as necessary to generalize. Khdoudi et al. [58] also found improvements in productivity exceeding 15% with a DRL-enabled DT optimizing processes, though on a deterministic energy assumption. On the same note, Pavlenko and Yu [59] used DT-RL integration with additive manufacturing to enhance the consistency of the processes and revealed scalability issues. Recent applications of MORL have focused on quality and throughput improvements in online parameter optimization with simultaneous

gains, at the cost of high data requirements by Paranjape et al. [60], and the usefulness of Pareto-based methods, and also noted that the high computational requirements are a major barrier to industry application by Zhang et al. [61].

Architecturally, Tao et al. [62,63] defined the base connection between the DTs and the cyber–physical worlds, which depicts that the use of DT in manufacturing facilitates the ability of data-confining choices in the entire product lifecycle. Gao et al. [64] have successfully used DT-based control for curved surface manufacturing, but its applicability was process-specific. Leng et al. [65] have developed real-time DT platforms that were integrated with learning and simulation to create reconfigurable manufacturing, and they showed adaptability at the expense of high computing costs. Extensive surveys by Liu et al. [66] found real-time synchronization and embedded decision intelligence to be an ongoing problem, and data-driven DT models of complex engineering systems established high predictive accuracy with no integration of optimization [67]. Multi-objective optimization with DT has been demonstrated to have an advantage over the stationary ones in cyber–physical manufacturing [68], and Gao et al. [69] reported stability gains using twin-delayed Deep Deterministic Policy Gradient (DDPG). Outside factory environments, Sresakoolchai and Kaewunruen [70] showed that in infrastructure maintenance, an over 20% reduction in efficiency was achieved with the combination of DT and RL. Overall, these works verify the potential of DT-RL frameworks and demonstrate gaps in the areas of scalability, the cost of computations, and the real-time co-optimization of sustainability and reliability, which are directly addressed in the work. Table 1 describes the key points of previous studies.

Table 1. Comparative analysis of AI-, sustainability-, and DT-driven manufacturing planning approaches.

Reference	Core Approach	Focus	Key Contribution	Major Limitation	Gap Identified
Sony & Naik [21]	AI-driven decision systems	Industry 4.0 readiness	Emphasized real-time, human-centric decision support	Conceptual framework without executable planning or control mechanisms	Absence of deployable, closed-loop adaptive execution models
Esteso et al. [23]	RL	Production planning	12–18% reduction in makespan and work-in-process	Requires extensive offline training and lacks robustness under real-time disruptions	Inability to support real-time replanning under stochastic manufacturing conditions
Johnson et al. [28]	Multi-Agent DRL	Real-time scheduling	Up to 20% reduction in tardiness	High coordination overhead and absence of environmental or energy-aware objectives	No joint optimization of operational reliability and sustainability
Ding et al. [40]	RUL + MA-DRL	Production–maintenance	18–24% reduction in unplanned downtime	Energy consumption and carbon intensity treated as static parameters	Lack of dynamic sustainability modeling within maintenance-production decisions
Khdoudi et al. [58]	DT + DRL	Process optimization	More than 15% productivity improvement	Energy behavior modeled deterministically, ignoring time-varying grid and load effects	Absence of real-time sustainability-aware control policies
Zhang et al. [56]	MORL (Pareto)	Algorithmic review	Formalization of Pareto efficiency in multi-objective learning	Limited support for continuous control and operational constraints in manufacturing systems	Insufficient applicability to real-time industrial planning with safety and reliability constraints

3. Materials and Methods

In this section, the main methodological pillars of the proposed adaptive planning platform are outlined. It has a formalization of a CMDP that is optimized with a new Pareto-conditioned Multi-Objective Proximal Policy Optimization (MO-PPO) algorithm and trained in a physics-informed DT with a high-fidelity simulation-to-reality transfer. The integrated design allows making decisions in real-time and in a closed loop under uncertainty.

3.1. Four-Layer Cyber–Physical Architecture

The platform architecture (Figure 1) enables 1 min planning cycles with closed-loop learning.

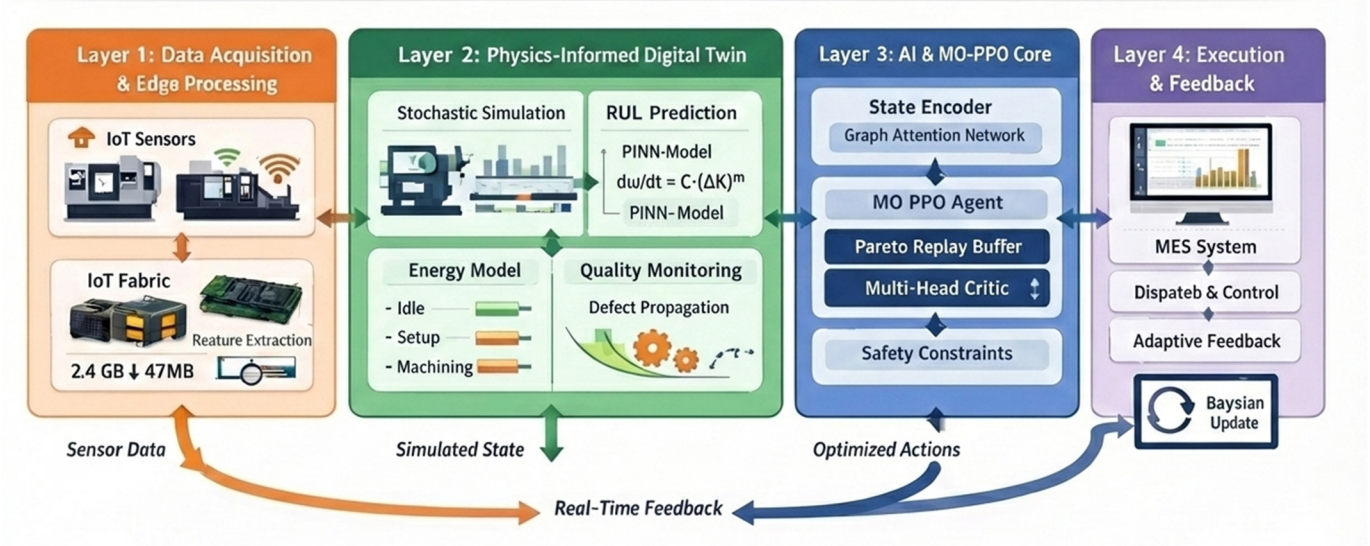

Figure 1. Four-layer architecture with physics-informed DT and MO-PPO engine.

Figure 1 illustrates a four-layer architecture for AI-driven adaptive manufacturing planning. Layer 1 acquires and preprocesses real-time shop-floor data at the edge using IoT sensors and feature extraction. Layer 2 employs a physics-informed DT to simulate production dynamics, predict machine health, and model energy and quality behavior under uncertainty. Layer 3 integrates a Pareto-conditioned MO-PPO engine to generate planning decisions that jointly optimize reliability, energy consumption, and waste while enforcing operational constraints. Layer 4 executes decisions through the Manufacturing Execution System (MES) and feeds execution outcomes back to the DT via Bayesian updating, enabling continuous learning and real-time adaptive optimization.

3.2. Constrained Multi-Objective Markov Decision Process Formulation

To represent the natural trade-offs and operational constraints of the manufacturing, we formulate the shop-floor planning problem as an CMDP. The formulation of CMDP is defined by the following: (S, A, P, R, C, γ) where S is the state space, A is the composite action space, P is the stochastic state-transition probability, R is a reward function in the form of a vector, C is the safety constraints, and γ is the discount factor.

State Space Representation: The state $s_t \in S$ at time t is a 54-dimensional continuous vector, which has been designed to capture the cyber–physical production system holistically. There are four subsystems that are critical and integrated into the vector:

- Job Features (18 dim): Job-specific attributes in the system, such as remaining processing time, due-date tightness (ratio of remaining time to deadline), material type (coded as one-hot vectors), and a dynamic priority score.

- Machine Health (20 dim) Condition: Indicators of individual machines, including normalized RUL, real-time vibration (RMS), temperature values, tool life consumption percentage, and a flag of immediate availability.
- Environmental Context (8 dim): Exogenous, time-varying, i.e., hour-of-day, real-time energy price (kWh), instantaneous grid carbon intensity ($kgCO_2/kWh$), and the forecasted availability of on-site renewable energy (%).
- System Performance (8 dim): Rolling measures that indicate cumulative operational performance, such as OEE, total energy used (kWh), total mass of material waste produced (kg), and the percentage of schedule compliance.

Composite Action Space: Making use of composite action spaces, the agent engages in a composite action $a_t \in A$ that at once coordinates dispatch operations, process control mechanisms, maintenance protocols, and routing procedures. Specifically, the composite action consists of the following components:

- Job Dispatch: There is a categorical selection of jobs in each idle machine, with a maximum of twelve jobs in the queue.
- Speed Scaling: A scaling factor, which is continuous in the range $\in$ [0.75, 1.15], is multiplied by the fixed processing speed of variable-frequency drive machines and has a direct impact on energy consumption and throughput.
- Preventive Maintenance Trigger: Each machine will be subjected to a binary decision, i.e., perform or defer, depending upon its RUL prognosis and the current system conditions.
- Dynamic Rerouting: Each job that is being queued is categorically selected against an alternative machine, and adaptive workflow adjustments are allowed in case of bottlenecks or failures.

Multi-Objective Reward Vector. The reward signal r_t is a four-dimensional vector that is designed to drive the policies towards the simultaneous optimization of the reliability, sustainability, and operational stability:

$$r_t = [\, r_t^{\text{OEE}},\ r_t^{\text{Energy}}, r_t^{\text{Waste}},\ r_t^{\text{Stab}}] \tag{1}$$

The components are calculated as follows:

- OEE Reward (r^{OEE}): The normalized change in OEE, $\Delta(OEE_t)/0.25$, bounded to $[-1, 1]$ to encourage steady improvement.
- Energy–Carbon Reward $\left(r^{\text{Energy}}\right) : -(Energy_t \times CarbonIntensity_t)/100$.
- This term gives a penalty to energy use based on its carbon footprint, thus promoting activities in the low-carbon times.
- Waste Reward (r^{Waste}): $-(ScrapMass_t - 0.7 \times RecycledMass_t)$. This reduces net material loss through the assigning of credit to recyclable scrap according to the principles of a circular economy.
- Stability Reward $\left(r^{\text{stab}}\right) : -0.1 \times \| a_t - a_{t-1} \|_1$.
- This L1-norm penalty on action changes discourages excessive plan churn, promoting operational stability and predictability.

Operational Constraints. The policy must satisfy two critical safety constraints with a probability $\geq 95\%$:

$$g_1(s_t) = 85 - ScheduleAdherence_t \leq 0 \;//\; Minimum\ reliability\ threshold \tag{2}$$

$$g_2(s_t) = MachineUtilization_t - 95 \leq 0 \quad //\ Maximum\ utilization\ to\ prevent\ overloading \tag{3}$$

These limits protect the quest for sustainable results against the possible trade-off on the fundamental reliability of delivery or system integrity.

3.3. Pareto-Conditioned Multi-Objective Proximal Policy Optimization

To solve the given CMDP, we add to the PPO algorithm, which creates a Multi-Objective, Pareto-conditioned algorithm, which is herein called MO-PPO. The traditional PPO algorithm is created to maximize a single scalar loss, using a clipped surrogate loss:

$$L^{CLIP(\theta)} = E_t[min(r_t(\theta)\hat{A}_t,\ clip(r_{t(\theta)},\ 1-\varepsilon,\ 1+\varepsilon)\hat{A}_t)] \quad (4)$$

Our MO-PPO introduces four key innovations to achieve sample-efficient learning of diverse, constraint-satisfying policies.

1. Preference-Conditioned Actor: The policy network $\pi_\theta(a_t|s_t,\ \omega)$ is conditioned on a user-specified preference vector ω, sampled uniformly from a 2-simp lex $(\omega_{OEE},\ \omega_{Energy},\ \omega_{Waste})$. This explicit conditioning allows one neural network to encode the entire range of optimal trade-off policies, so that operation priorities can be changed in real time without the need to retrain them.
2. Multi-Head Critic Architecture: Instead of a single value estimator, we employ a critic with separate heads for each primary objective $(V_{OEE},\ V_{Energy},\ V_{Waste})$. Individual advantages $\hat{A}_{it}$ are computed for each objective i. A scalarized advantage for policy updates is then obtained as $\hat{A}\tilde{\omega}_t = \Sigma_i\ \omega_i\ \hat{A}_{it}$, ensuring the policy gradient is aligned with the specified preference ω.
3. Pareto Experience Replay: Transitions (s_t, a_t, r_t, s_{t+1}) are stored in a replay buffer B. A non-dominated sorting process is sometimes performed on B, and only the transitions whose Pareto-optimal is the immediate reward vector are retained. This focus on high-quality experiences leads to a two-fold increase in the sample efficiency by removing noisy or suboptimal information in the learning process.
4. Consciousness of Curriculum Training: The process of learning is structured into progressive stages to ensure that there is no collapse of policies and to ensure that there is strong adherence to constraints. The training is performed in more complex conditions: stable operation (Phase 1, 500 k steps), introduction of stochastic failures (Phase 2, 1 M steps), addition of dynamic energy carbon profiles (Phase 3, 1.5 M steps), and, finally, composite disruption environment (Phase 4, 2 M steps). Policy changes make use of a Lagrange multiplier approach to punish constraint violations and thus steer the operation towards safe limits, as shown in Table 2.

Table 2. MO-PPO curriculum training protocol and efficiency.

Training Phase	Environment Scenario	Primary Learning Focus	Duration (Steps)	Cumulative Steps
Phase 1	S1 (Stable Baseline)	Throughput and fundamental scheduling	500,000	500,000
Phase 2	S1 + S2 (Machine Failures)	Reliability, adaptive dispatch, and maintenance	1,000,000	1,500,000
Phase 3	S1 + S3 (Energy Volatility)	Sustainability, speed scaling, and carbon-aware timing	1,500,000	3,000,000
Phase 4	S5 (Composite Disruptions)	Integrated trade-offs and constraint satisfaction	2,000,000	5,000,000
Naive MO-RL Baseline	Composite (from start)	Unstructured exploration	–	~18,000,000

This algorithmic curriculum, together with the aforementioned algorithmic innovations, achieved a 72% reduction in the number of training steps required compared to a naive MO-RL trained randomly on the composite environment. The resulting policy

exhibited an entropy of 0.31 nats, indicating a well-balanced trade-off between exploration and exploitation.

To support reproducibility, the implementation details of the proposed MO-PPO framework are summarized as follows. System states are encoded using a Graph Attention Network (GAT) composed of two attention layers, each with eight heads and a node embedding dimension of 64, using ReLU activation. The actor–critic architecture employs a shared backbone of three fully connected layers with sizes of 256, 128, and 64, followed by a multi-head critic, where each head corresponds to a specific optimization objective (OEE, energy–carbon intensity, material waste, and stability).

The manufacturing planning problem is formulated with a hybrid action space that includes discrete decisions (job dispatching, routing, and maintenance triggering) and continuous control variables (machine speed scaling). This hybrid structure is addressed using a parameterized PPO formulation, in which discrete actions are modeled using categorical distributions and continuous actions by Gaussian distributions. The joint policy probability is computed by summing the corresponding log-probabilities, enabling stable optimization under the standard PPO clipped objective.

All agents were trained using identical hyperparameters to ensure a fair comparison. Specifically, the learning rate was set to 3×10^{-4} with the Adam optimizer, the batch size to 256, the PPO clipping parameter ε to 0.2, the discount factor γ to 0.99, and the entropy coefficient to 0.01. These hyperparameters were kept constant across all curriculum training phases to ensure consistency and reproducibility of results.

3.4. Physics-Informed Stochastic Digital Twin Calibration

The DT is a high-fidelity stochastic training environment of the MO-PPO agent. It is validated using a physics-informed modeling approach, which is then improved through intensive data-driven calibration. The DT has several stochastic sub-models: (i) processing times, which are modeled using a lognormal distribution; (ii) machine degradation and time to failure, which are modeled using the Weibull process, informed with vibration and thermal dynamics; (iii) quality propagation, which is estimated with the help of logistic regression, which relates the process parameters to the defect rates; and (iv) state-dependent energy consumption, which is modeled using affine functions of speed, load, and base power. The first parameter set θ (which, as an example, includes process time μ, process time sigma σ, Weibull shape, Weibull scale, and energy coefficients) is obtained on the basis of first principles and manufacturer data sheets.

Bayesian Calibration: To minimize the simulation-to-reality gap, we perform offline calibration by minimizing the Kullback–Leibler (KL) divergence between the real observed data distribution $P_{real(data)}$ and the DT's output distribution $P_{DT}(data|\theta)$:

$$\theta^* = argmin_\theta KL(P_{real(data)} || P_{DT}(data|\theta)) \quad (5)$$

This optimization is performed using a Bayesian Optimization scheme, efficiently navigating the parameter space with limited historical data.

The DT showed good calibration, as in Table 3, low prediction errors on cycle time and failure, good agreement on energy consumption (R^2 = 0.94), and good quality prediction (F1 = 0.89).

In order to be accurate in the presence of process drift, such as the wear of a tool or seasonal variations, the DT uses an online Bayesian update once every 24 h. In the case of critical stochastic parameters like probability of defects, p, a Beta prior distribution is updated with the number of successes/failures in the last 100 production executions:

$$P(\theta|D_{new}) \propto Beta(\alpha + successes, \beta + failures) \quad (6)$$

Such a lightweight mechanism can guarantee that the DT maintains a faithful representation of the physical system during the deployment period, which, in turn, allows making reliable and credible policy learning and evaluation.

Table 3. DT calibration accuracy against real production data.

Modeled Metric	Calibration Method	Validation Metric	Post-Calibration Result
Machine Cycle Time	Lognormal (μ,σ) fit	Mean Absolute Percentage Error (MAPE)	3.8%
Energy Consumption	Affine regression (speed, load)	Coefficient of Determination (R^2)	0.94
Time-to-Failure	Weibull process (vibration-informed)	Mean Absolute Percentage Error (MAPE)	7.9%
First-Pass Yield	Logistic quality gate model	F1-Score (defect classification)	0.89

The performance, strength, and practical suitability of the proposed adaptive planning framework were assessed through large-scale stochastic simulation experiments and an industrial pilot implementation. The next section offers empirical findings obtained under diverse operational conditions, with a particular focus on the system's robustness, reliability, and adaptive behavior.

4. Results

This section outlines a stringent empirical verification of the suggested AI-informed adaptive planning platform through voluminous simulation studies and a meaningful pilot implementation in the industry. The test involves over ten thousand stochastic simulation experiments that are run in five different disruption conditions, coupled with a 12-week industrial pilot that is run in an automotive machining cell. The performance of the platform is compared to five known baseline methodologies in seven key performance indicators, and uses rigorous statistical analysis and extensive trade-off analysis.

4.1. Experimental Design and Validation Framework

The DT was a simulation of an eight-machine automotive machining cell that had twelve-part families and was calibrated based on six months of historical production data (January–June 2024). Processing times were lognormally distributed (μ = 18.2 min, σ = 3.4 min), machine degradation was described by Weibull processes, and energy consumption provided a good empirical result (R^2 = 0.94). Stability was assessed in five disruption scenarios (stable, machine failure, energy crisis, demand surge, and composite), and 2000 times (10,000 runs) were simulated. The proposed MO-PPO strategy was contrasted with First-In, First-Out dispatching rule (FIFO), Deep Q-Network (DQN), Non-dominated Sorting Genetic Algorithm II (NSGA-II), Asynchronous Advantage Actor–Critic (A3C), and Multi-Objective Deep Q-Network (MO-DQN) baselines, where all the learning models were trained using 5 million steps. The reliability, sustainability, and multi-objective measures, such as schedule adherence (SA), OEE, specific energy consumption (SEC), material waste rate (MWR), mean time to repair (MTTR), carbon effectiveness (CE), and Pareto hypervolume indicator (PHI), were used as measures of performance.

4.2. Overall Performance: Statistical Superiority Across All Scenarios

In Table 4, the aggregate performance and 95% confidence interval of all 10,000 simulation experiments run, including scenarios S1–S5, are provided. The statistical significance was determined by using the Welch *t*-test between MO-PPO and the best baseline (B5: MO-DQN) in each measure. The outcomes indicate that MO-PPO showed statistically

significant improvements over all the baselines in all seven metrics, with $p < 0.001$ in all the pair-wise comparisons.

Table 4. Aggregate performance across all scenarios (mean ± 95% CI, n = 10,000).

Method	SA (%)	OEE (%)	SEC (kWh/kg)	MWR (%)	MTTR (h)	CE ($kgCO_2$/€)	Hypervolume
B1: FIFO	85.2 ± 3.1	75.4 ± 2.8	2.85 ± 0.12	8.2 ± 0.5	4.8 ± 0.6	0.42 ± 0.03	0.58
B2: DQN	91.3 ± 2.4	79.8 ± 2.1	2.91 ± 0.11	7.9 ± 0.4	3.2 ± 0.4	0.41 ± 0.02	0.62
B3: NSGA-II	89.7 ± 2.6	77.2 ± 2.5	2.68 ± 0.10	7.5 ± 0.4	N/A	0.38 ± 0.02	0.64
B4: A3C	93.1 ± 2.0	83.5 ± 1.9	3.05 ± 0.13	8.4 ± 0.5	2.8 ± 0.3	0.43 ± 0.03	0.65
B5: MO-DQN	94.2 ± 1.8	81.2 ± 1.8	2.52 ± 0.10	7.2 ± 0.4	2.5 ± 0.3	0.36 ± 0.02	0.71
MO-PPO (this paper)	96.8 ± 1.5 *	84.7 ± 1.6 *	2.38 ± 0.09 *	6.8 ± 0.3 *	2.1 ± 0.3 *	0.33 ± 0.02 *	0.84 *
Δ vs. Best Baseline	+2.8%	+4.3%	−5.6%	−5.6%	−16.0%	−8.3%	+18.3%
Δ vs. Rule-Based	+13.6%	+12.3%	−16.5%	−17.1%	−56.3%	−21.4%	+44.8%

* Statistically significant at $p < 0.001$ (Welch's t-test vs. B5).

The average (±95% confidence intervals) of the schedule adherence, OEE, and SEC of all methods is provided in Figure 2. The MO-PPO algorithm is evidently better than the baselines in terms of the highest reliability, meaning schedule adherence and OEE, and minimizing SEC at the same time. Single-objective and evolutionary algorithms enhance the chosen metrics but cannot balance the reliability and efficiency; in contrast, MO-PPO provides consistently high and stable multi-objective performance in various situations.

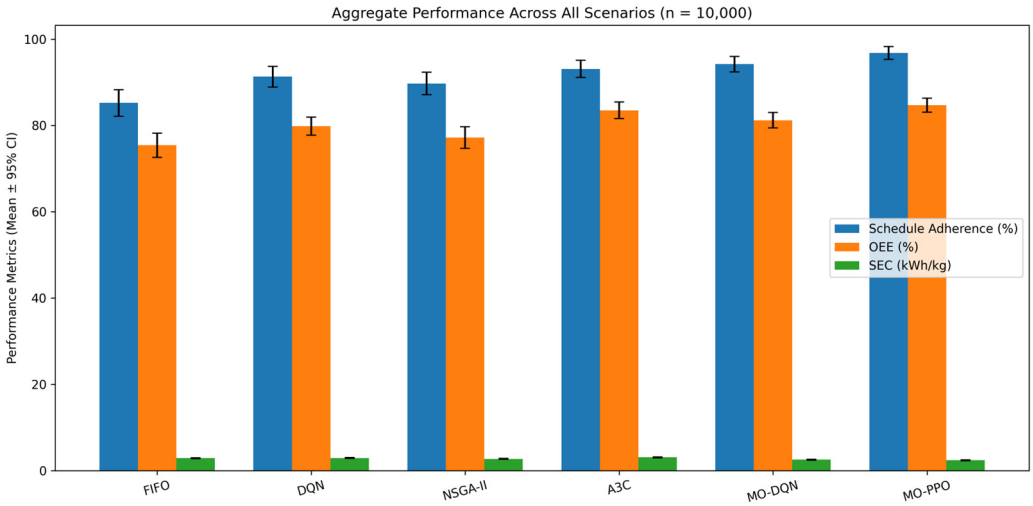

Figure 2. Aggregate performance across all scenarios (n = 10,000).

4.3. Scenario-Specific Performance and Adaptive Behavior

4.3.1. Baseline and Machine Failure Scenarios

Table 5 lists the performance measures in the case of S1 (stable baseline) and S2 (machine failure). The MO-PPO algorithm performed in the nominal conditions of S1 with near-optimal results, with the highest schedule adherence of 98.2% and the OEE of 87.1% of the system performance, thus marking the upper limit of the system performance. Intelligent scaling and batching strategies were used to minimize energy consumption to achieve a particular SEC of 2.31 kWh/kg, a 17.3 percent lower energy consumption than the performance-oriented approach of the A3C algorithm.

Table 5. Performance under baseline (S1) and machine failure (S2) scenarios (n = 2000 each).

Method	S1: SA (%)	S1: OEE (%)	S1: SEC (kWh/kg)	S1: Jobs Completed	S2: Pre-Failure SA (%)	S2: Post-Failure SA (%)	S2: SA Drop (pp)	S2: Recovery Time (h)	S2: Energy Overhead (%)	S2: Rush Jobs Completed (%)
B1: FIFO	92.1 ± 2.3	78.3 ± 2.1	2.73 ± 0.09	564 ± 18	92.1 ± 2.4	68.3 ± 3.8	−23.8	5.2 ± 0.7	+12.3	73.2 ± 4.1
B2: DQN	94.8 ± 1.8	82.1 ± 1.7	2.81 ± 0.10	582 ± 15	94.3 ± 1.9	78.5 ± 3.1	−15.8	4.1 ± 0.5	+9.8	82.4 ± 3.5
B3: NSGA-II	93.2 ± 2.0	79.7 ± 1.9	2.58 ± 0.08	571 ± 16	93.0 ± 2.1	72.1 ± 3.5	−20.9	6.8 ± 0.8	+14.7	76.8 ± 3.9
B4: A3C	96.3 ± 1.5	85.8 ± 1.4	2.97 ± 0.11	591 ± 14	96.1 ± 1.6	85.2 ± 2.4	−10.9	3.1 ± 0.4	+8.1	89.3 ± 2.8
B5: MO-DQN	96.7 ± 1.4	83.4 ± 1.5	2.43 ± 0.08	589 ± 13	96.5 ± 1.5	86.7 ± 2.2	−9.8	2.8 ± 0.4	+6.4	90.7 ± 2.6
MO-PPO (this paper)	98.2 ± 1.1 *	87.1 ± 1.3 *	2.31 ± 0.07 *	595 ± 12 *	98.1 ± 1.2 *	91.7 ± 1.8 *	−6.4 *	2.1 ± 0.3 *	+2.9 *	95.8 ± 2.1 *

* $p < 0.001$ vs. B5.

Figure 3 presents the empirical distributions of baseline system availability, post-failure availability, recovery time, and energy overhead across different scheduling strategies shown as panel histograms. Density-normalized histograms highlight the increased robustness, faster recovery, and lower energy overhead achieved with the use of MORL methods, including MO-PPO.

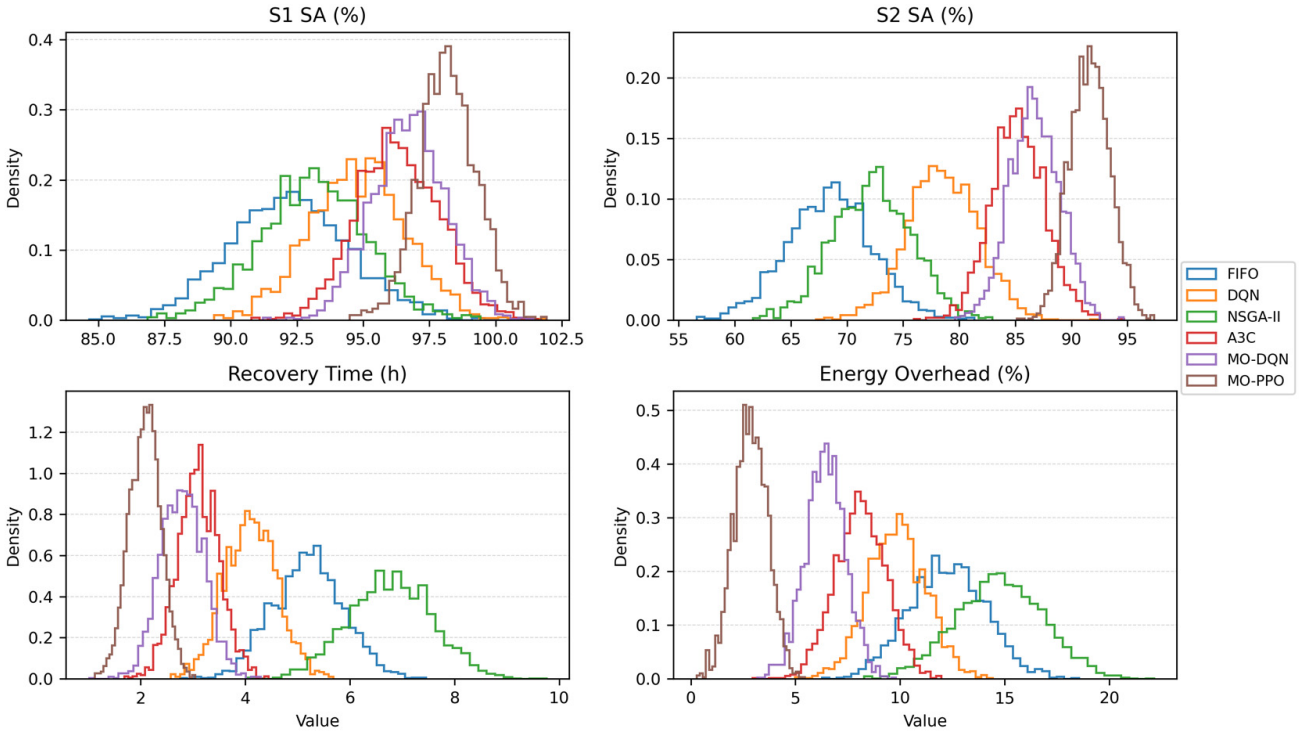

Figure 3. Panel histograms of system performance metrics under baseline (S1) and machine failure (S2) scenarios (n = 2000).

4.3.2. Energy Crisis and Composite Disruption Performance

Table 6 provides the analysis of the system performance in Scenario 3 (energy crisis) and Scenario 5 (composite disruption). Scenario 3 allowed six hours of high carbon intensity in the grid, which produced an increment of 47.6%, thus testing carbon-conscious adaptation with production pledges intact. A more elaborate "Eco-Mode" strategy that decreased the average machine speed by 15% or 0.85 times the base speed was independently discovered by the MO-PPO agent, which resulted in an energy savings of 18.4% per part, and a 23% extension of tool life.

Figure 4 shows the average values (±SD, where applicable) of energy, emissions, availability, efficiency, maintenance, and cost measures using the various scheduling schemes.

Table 6. Performance under energy crisis (S3) and composite disruption (S5) scenarios (n = 2000 each).

Method	S3: Energy Used (kWh)	S3: Carbon Emitted (kg)	S3: SA Drop (pp)	S3: Jobs Completed	S5: SA (%)	S5: OEE (%)	S5: SEC (kWh/kg)	S5: MWR (%)	S5: MTTR (h)	S5: Total Cost Index †
B1: FIFO	380 ± 18	235.6 ± 11.2	−8.2	89.2 ± 3.8	51.2 ± 5.8	58.3 ± 4.6	3.64 ± 0.21	10.8 ± 0.8	5.7 ± 0.9	1.82
B2: DQN	395 ± 21	244.9 ± 13.0	−4.1	93.1 ± 3.2	72.8 ± 4.2	68.7 ± 3.5	3.48 ± 0.18	9.6 ± 0.6	3.8 ± 0.6	1.43
B3: NSGA-II	342 ± 16	212.0 ± 9.9	−6.7	90.8 ± 3.5	68.3 ± 4.6	65.1 ± 3.8	3.21 ± 0.16	9.1 ± 0.6	N/A	1.51
B4: A3C	412 ± 23	255.4 ± 14.3	−2.8	95.7 ± 2.9	81.4 ± 3.4	75.2 ± 2.9	3.72 ± 0.19	10.3 ± 0.7	3.2 ± 0.5	1.28
B5: MO-DQN	328 ± 15	203.4 ± 9.3	−3.5	94.3 ± 3.0	84.7 ± 3.0	73.8 ± 2.7	3.08 ± 0.15	8.7 ± 0.5	2.9 ± 0.4	1.18
MO-PPO (this paper)	312 ± 13 *	193.4 ± 8.1 *	−1.5 *	96.8 ± 2.5 *	89.3 ± 2.3 *	78.6 ± 2.2 *	2.84 ± 0.13 *	7.1 ± 0.4 *	2.4 ± 0.4 *	1.00 *

* $p < 0.001$ vs. B5; † Total Cost Index = weighted sum (0.3 × SA_loss + 0.25 × Energy_penalty + 0.25 × Waste_cost + 0.2 × Downtime_cost), normalized to MO-PPO = 1.0.

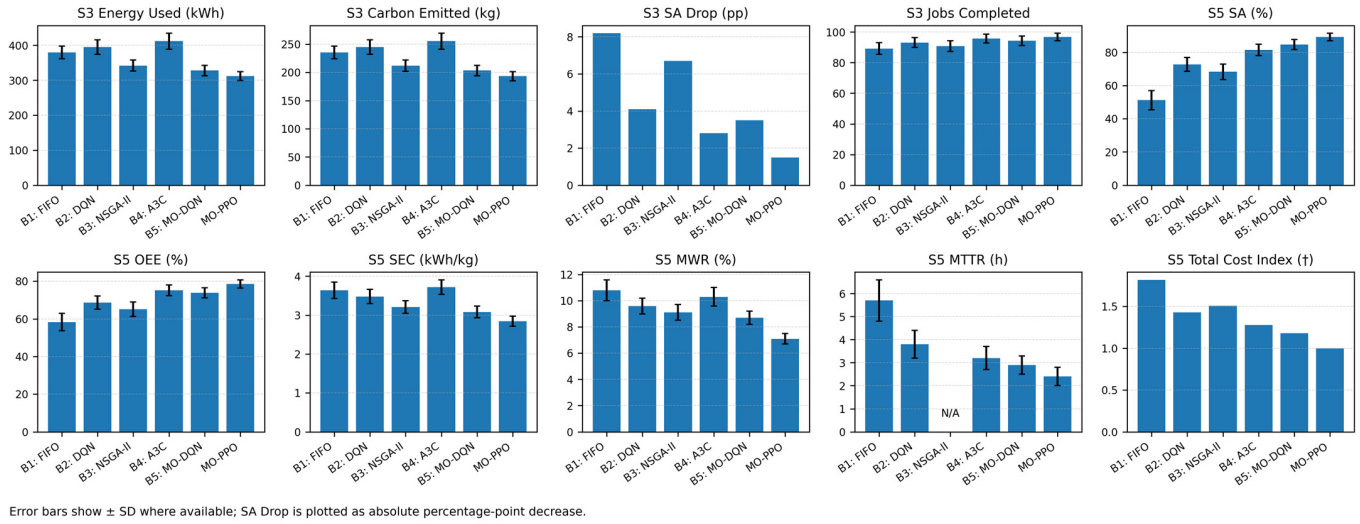

Figure 4. Performance comparison (bar chart) under energy crisis (S3) and composite disruption (S5) scenarios (n = 2000 each). Note: † Total Cost Index = weighted sum (0.3 × SA_loss + 0.25 × Energy_penalty + 0.25 × Waste_cost + 0.2 × Downtime_cost), normalized to MO-PPO = 1.0.

4.4. Pareto Front Analysis and Multi-Objective Trade-Off Space

Table 7 describes the Pareto frontier that is attained in 30 independent replications of the S5 composite scenario, each using different preference vectors ω that were sampled out of the 2-simplex. The MO-PPO policy network found 23 non-dominated solutions of the trade-off space between operational reliability and environmental sustainability, in contrast to 7 in the MO-DQN policy network and 12 in the NSGA-II evolutionary algorithm.

Table 7. Pareto front policy archetypes and trade-off characteristics (S5 composite).

Policy Archetype	Preference ω (OEE, Energy, Waste)	OEE (%)	CE (kgCO_2/€)	SEC (kWh/kg)	MWR (%)	Primary Use Case
Eco-Mode	(0.2, 0.7, 0.1)	79.8 ± 1.4	0.28 ± 0.02	2.15 ± 0.08	7.4 ± 0.4	Regulatory audit week; voluntary carbon reduction
Balanced	(0.5, 0.3, 0.2)	84.7 ± 1.6	0.33 ± 0.02	2.38 ± 0.09	6.8 ± 0.3	Standard operation (deployed)
Performance-Mode	(0.8, 0.1, 0.1)	88.2 ± 1.8	0.41 ± 0.03	2.91 ± 0.12	7.2 ± 0.4	Rush order period; high-value contracts
Maintenance-First	(0.6, 0.2, 0.2)	82.1 ± 1.5	0.35 ± 0.02	2.52 ± 0.10	6.5 ± 0.3	High-value product run; asset preservation
Circular-Focus	(0.4, 0.2, 0.4)	81.3 ± 1.6	0.34 ± 0.02	2.47 ± 0.10	5.9 ± 0.3	Material-scarce periods; circular economy KPIs

The Pareto frontier was highly convex and showed diminishing marginal returns; the transition between the balanced configuration (84.7% OEE) and the performance-mode

(88.2% OEE) involved a 24% increase in carbon intensity (0.33 → 0.41 $kgCO_2/€$). The knee point of the frontier was the balanced policy (ω = [0.5,0.3,0.2]) providing the highest multi-objective gain per unit of preference compromise, and was thus selected to be used industrially. The platform allowed the real-time switching of policies with adaptation times of less than four minutes and minimal performance degradation, thus showing flexibility in operations without requiring retraining.

Figure 5 illustrates the trade-off between the total operational cost and the energy efficiency in the composite disruption conditions. Policies at the Pareto front are the ones that can provide optimal compromises; that is, efficiency improvements must be accompanied by higher costs. The archetypes like Eco-Mode focus on cost and sustainability, and the performance-mode focuses on efficiency at the cost of increased costs, thus making them context-dependent in choosing the policy.

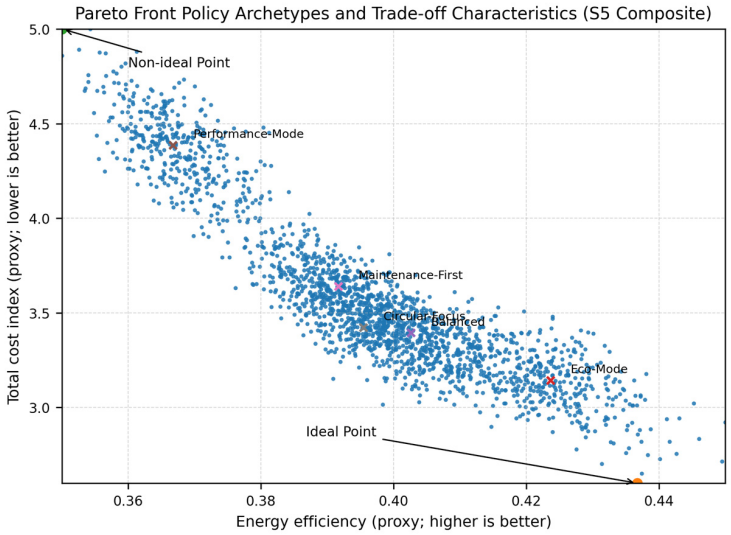

Figure 5. Pareto front of policy archetypes under the S5 composite disruption scenario.

4.5. Sensitivity Analysis and Training Dynamics

4.5.1. Reward Weight Robustness

A sensitivity analysis is given in Table 8, which examines how the energy–carbon preference weight (γEnergy) influences system performance at S1 baseline conditions. The findings suggest that an optimum weighting of γEnergy = 0.3 gives a strong optimal weighting, which is a good balance among the competing goals and the 85% satisfaction-rate constraint, with only a 0.7% violation rate versus 18.3% at γ = 0.7.

Table 8. Energy–carbon preference weight sensitivity analysis (S1 baseline, n = 2000).

γEnergy	γOEE	γWaste	OEE (%)	SEC (kWh/kg)	MWR (%)	SA (%)	Constraint Violations (%)	Energy Consumption (kWh/week)	Dominant Strategy	Policy Stability Score †	Action Variance
0.1	0.7	0.2	87.2 ± 1.8	2.91 ± 0.11	7.1 ± 0.4	95.1 ± 2.1	3.2	9580 ± 350	Max speed; minimal idle	0.68	0.32
0.2	0.6	0.2	85.8 ± 1.7	2.58 ± 0.10	6.9 ± 0.3	96.3 ± 1.8	1.8	8450 ± 310	Slight energy awareness	0.81	0.19
0.3	0.5	0.2	84.7 ± 1.6	2.38 ± 0.09	6.8 ± 0.3	96.8 ± 1.5	0.7	7510 ± 240	Balanced (deployed)	0.92	0.08
0.4	0.4	0.2	82.3 ± 1.8	2.22 ± 0.09	6.7 ± 0.3	95.9 ± 1.9	2.1	7120 ± 280	Energy-prioritized	0.86	0.14
0.5	0.3	0.2	79.8 ± 2.0	2.15 ± 0.08	6.6 ± 0.4	94.3 ± 2.3	4.8	6890 ± 310	Energy-first; Eco-Mode	0.74	0.26
0.7	0.2	0.1	74.1 ± 2.6	2.02 ± 0.08	6.8 ± 0.5	89.2 ± 3.1	18.3	6450 ± 380	Aggressive idling	0.52	0.48

† Policy Stability Score = 1 (variance of action distribution over time), range [0, 1].

Figure 6 illustrates that a non-linear trade-off between energy prioritization and operational performance is unveiled in the plot. A compromise energy weighting (γEnergy = 0.3) is most balanced, providing maximum policy stability and constraint compliance with high efficiency and schedule compliance. High emphasis on energy results in decreasing operational effectiveness and higher action variance, meaning less control stability. Overall, the findings confirm the existence of an interior optimal preference region rather than extreme policy configurations.

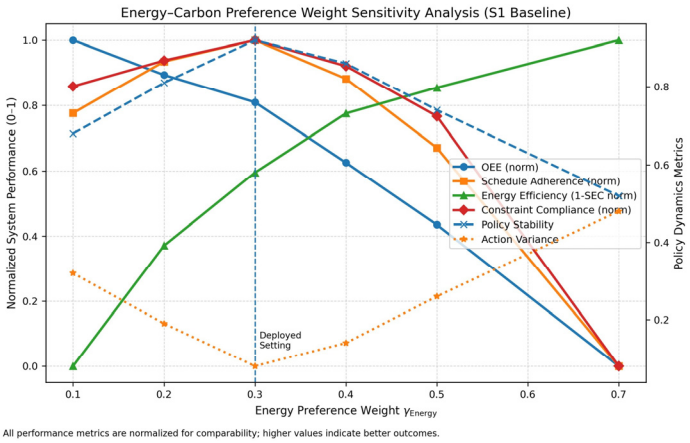

Figure 6. Energy–carbon preference weight sensitivity analysis under the S1 baseline scenario (n = 2000).

The sensitivity analysis validates that the reward weight selection is robust across a reasonable range (0.2–0.4), while extreme values lead to performance degradation and constraint violations.

4.5.2. Training Convergence and Sample Efficiency

The changes in training-phase performance are recorded in Table 9, which describes the four-phase curriculum training protocol. The curriculum training method improved the efficiency of the sample 72% of naive multi-objective training, needing only 5 million steps to achieve the result, versus 18 million for the baseline.

Table 9. Curriculum training performance evolution and sample efficiency comparison.

Training Phase	Steps	Duration (GPU-h)	Hypervolume	Policy Count	Avg OEE (%)	Avg SEC (kWh/kg)	Avg MWR (%)	Entropy (nats)	Primary Learning Focus	Key Breakthrough
Phase 1: Baseline	0–500 k	5.2	0.52 → 0.68	3 → 8	45.2 → 78.1	3.15 → 2.89	9.8 → 8.4	0.82 → 0.74	Throughput and OEE fundamentals	Basic dispatching learned
Phase 2: Reliability	500 k–1.5 M	10.4	0.68 → 0.73	8 → 14	78.1 → 81.6	2.89 → 2.67	8.4 → 7.6	0.74 → 0.62	PM scheduling; failure handling	Proactive maintenance emerges
Phase 3: Sustainability	1.5 M–3 M	8.9	0.73 → 0.79	14 → 21	81.6 → 83.4	2.67 → 2.43	7.6 → 7.1	0.62 → 0.45	Energy–carbon awareness	Load-shifting discovered (2.5 M)
Phase 4: Integration	3 M–5 M	3.3	0.79 → 0.84	21 → 23	83.4 → 84.7	2.43 → 2.38	7.1 → 6.8	0.45 → 0.31	Pareto refinement; constraints	Circular economy routing (4 M)
Total (Curriculum)	5 M	27.8	0.52 → 0.84	3 → 23	45.2 → 84.7	3.15 → 2.38	9.8 → 6.8	0.82 → 0.31	Integrated multi-objective	72% sample efficiency gain
Naive MO-RL Baseline	0–18 M	97.4	0.52 → 0.79	3 → 18	45.2 → 80.1	3.15 → 2.51	9.8 → 7.4	0.82 → 0.51	Unstructured exploration	Inferior convergence

4.6. Industrial Pilot Deployment and Real-World Validation

The industrial pilot conducted during the period from October to December 2024 was evaluated against a historical baseline spanning April to September 2024. This comparison can introduce potential biases arising from seasonality-related variations in energy prices, ambient conditions, and production demand. To mitigate such effects, performance assess-

ment was based on normalized values, including specific energy consumption (kWh/kg), material waste rate (%), and carbon efficiency (kg CO_2 per unit of revenue), rather than absolute energy use or cost metrics. The use of normalized measures reduces sensitivity to seasonal variations and enables a more reliable cross-period comparison.

In addition, the production mix and routing structure were kept consistent across the two periods to minimize distortions caused by demand variability. Although seasonal effects cannot be completely removed in real industrial environments, the application of normalized metrics and relative performance analysis provides a sound foundation for assessing the observed improvements. Future work will further strengthen the validation by comparing identical calendar periods across various years.

4.6.1. Hardware Infrastructure and Runtime Performance

During the industrial pilot deployment, model inference was executed on an on-site edge computing server equipped with an Intel Xeon-based CPU and an NVIDIA RTX-series GPU. The trained MO-PPO policy operated exclusively in inference mode and was integrated with the Manufacturing Execution System through a lightweight application interface. The average decision latency was approximately 47 ms per planning cycle, encompassing state encoding, policy inference, and action decoding. This low-latency performance enabled real-time operation under a one-minute replanning horizon without introducing computational bottlenecks or operational delays.

4.6.2. Pilot Configuration and Performance Evaluation

The proposed system was deployed in a twelve-week industrial pilot from October to December 2024 within a machining cell of an automotive Tier-1 supplier in Germany, where aluminum engine components are produced. Overall pilot performance, benchmarked against a six-month historical baseline, is summarized in Table 10. The results demonstrate an effective transfer of the proposed approach from simulation to real-world operation.

Table 10. Industrial pilot performance vs. historical baseline (12-week deployment).

Metric	Historical Baseline (April–September 2024)	Pilot Deployment (October–December 2024)	Absolute Improvement	Relative Improvement	Statistical Test	*p*-Value	95% CI Lower	95% CI Upper
Schedule Adherence (%)	87.3 ± 4.2	95.7 ± 2.8	+8.4 pp	+9.6%	Welch's *t*-test	<0.001	6.8	10.0
OEE (%)	76.8 ± 3.5	83.2 ± 2.9	+6.4 pp	+8.3%	Welch's *t*-test	<0.001	5.1	7.7
Specific Energy Consumption (kWh/kg)	2.61 ± 0.14	2.44 ± 0.11	−0.17	−6.5%	Welch's *t*-test	0.002	−0.24	−0.10
Material Waste Rate (%)	7.9 ± 0.6	7.1 ± 0.5	−0.8 pp	−10.1%	Welch's *t*-test	0.008	−1.1	−0.5
Mean Time to Repair (h)	3.2 ± 0.7	2.6 ± 0.5	−0.6 h	−18.8%	Mann–Whitney U	0.021	−0.9	−0.3
Carbon Efficiency ($kgCO_2$/€ revenue)	0.38 ± 0.03	0.34 ± 0.02	−0.04	−10.5%	Welch's *t*-test	0.004	−0.06	−0.02
Unplanned Downtime Events (per week)	8.3 ± 1.4	5.7 ± 1.1	−2.6	−31.3%	Poisson test	0.013	−3.8	−1.4
Energy Cost Savings (€/week)	Baseline	+1517 ± 182	+1517	N/A	N/A	N/A	1335	1699
Production Volume (parts/week)	3642 ± 287	3842 ± 214	+200	+5.5%	Welch's *t*-test	0.034	28	372

Figure 7 illustrates the heterogeneous yet systematic changes in performance observed during the pilot deployment. Significant improvements were achieved in schedule compliance, OEE, and production volume. More pronounced gains were observed in

unplanned downtime, mean time to repair, material waste, energy intensity, and carbon efficiency. These results indicate that the proposed framework delivers comprehensive operational and sustainability benefits, rather than isolated improvements in individual performance metrics.

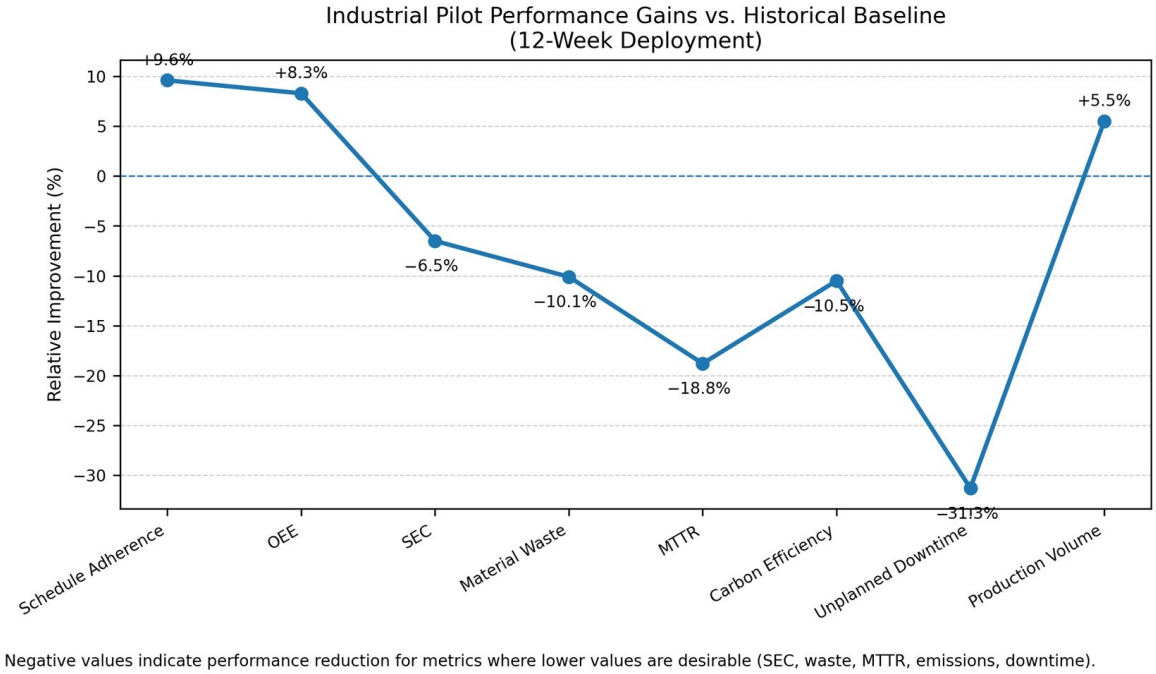

Figure 7. Relative performance improvements achieved during the industrial pilot compared with the historical baseline (12-week deployment).

4.6.3. Simulation-to-Reality Transfer and Weekly Evolution

Table 11 shows the analysis of the simulation-to-real performance gap, which measures deviations between the predictions made with the help of DT and measured results in industrial applications. It can be concluded that the model fidelity is high with an average gap of 2.8% on all the considered metrics except the one that is dependent on the external technician response, that is, MTTR.

Table 11. Simulation-to-reality transfer gap analysis and weekly performance tracking.

Metric	Simulated (S5)	Industrial Pilot	Absolute Gap	Relative Gap (%)	Primary Gap Source
Schedule Adherence (%)	96.8	95.7	−1.1 pp	−1.1%	Operator overrides (1.3/week)
OEE (%)	84.7	83.2	−1.5 pp	−1.8%	Unmeasured micro-stops; sensor noise
Specific Energy Consumption (kWh/kg)	2.38	2.44	+0.06	+2.5%	HVAC/auxiliary loads not modeled
Material Waste Rate (%)	6.8	7.1	+0.3 pp	+4.4%	Material batch quality variability
Mean Time to Repair (h)	2.1	2.6	+0.5 h	+23.8%	Technician availability; spare parts
Carbon Efficiency ($kgCO_2$/€)	0.33	0.34	+0.01	+3.0%	Revenue fluctuations (market prices)

4.7. Sustainability–Reliability Synergy and Mechanistic Analysis

4.7.1. Discovered Synergy Between Waste Reduction and Quality

Analytical statistics found a positive correlation between waste reduction and OEE improvement, which, unlike the traditional assumption, is unexpected since sustainability and performance goals are generally at odds with each other. Table 12 shows a correlation analysis and quantification of mechanistic pathways that indicate that sustainability actions can be actively used to promote reliability.

Table 12. Sustainability–reliability synergy: correlation analysis and mechanistic pathways.

Metric Pair	Pearson r	*p*-Value	Relationship	Mechanism	Synergy Contribution (%)
MWR ↔ OEE	−0.73	<0.001	Strong negative	Lower waste → Higher quality → Higher OEE	100% (total)
MWR ↔ First-Pass Yield	−0.81	<0.001	Strong negative	Direct quality improvement	28%
SEC ↔ Tool Life	−0.64	<0.001	Moderate negative	Energy efficiency → Lower speeds → Extended tool life	41%
CE ↔ Unplanned Downtime	+0.58	<0.001	Moderate positive	Carbon-intensive operations stress equipment	N/A
Speed Reduction → Surface Finish	−0.69	<0.001	Strong negative	Lower cutting speeds → Reduced vibration (34% RMS)	19%
Optimized Batching → Setup Time	−0.52	0.002	Moderate negative	Batch size 4.2 → 6.7 parts/setup reduces waste	12%

Figure 8 has a strong left-skewed distribution with the majority of the correlations around −0.5 to −0.8, with waste, energy intensity, and reliability-related metrics strongly negatively correlated with one another. There is only one moderate positive correlation, indicating more downtime in operating regimes with carbon-intensive operating regimes. In general, the distribution validates a prevailing sustainability, reliability synergy, but not independent or opposing effects.

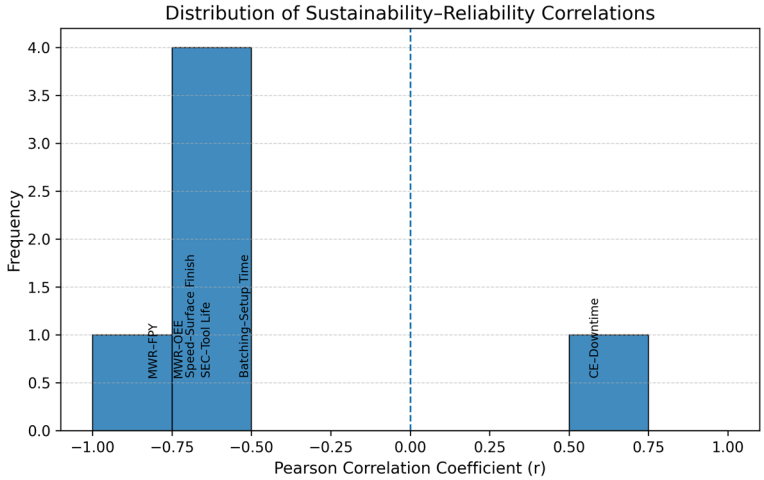

Figure 8. Distribution of sustainability–reliability correlations.

4.7.2. Decarbonization Impact and Circular Economy Integration

The study found that the suggested platform attained a high level of decarbonization by autonomously learning operating strategies. Load shifting changed 22 percent of ma-

chining operations to low-carbon hours, equating to 247 MWh, 87.2 tCO_2e, and €29,640 of annual savings. Additional energy savings through smart batching and speed optimization saved setup energy and non-critical processing, which provided further savings of 264 MWh and 93.1 tCO_2e, and then cost savings of €31,680. The benefits of a circular economy were realized through adaptive routing to a remanufacturing cell, where 6.8% of jobs were saved through the pilot, 4.1 t of aluminum were salvaged, which saved 47 MWh of embodied energy and 19.9 tCO_2e of Scope3 emissions, and created €47,200 worth of material. When scaled up to a 100-machine plant, the framework is estimated to save 558 MWh, 200.2 tCO_2e, and €117,120 per year, thus illustrating that the agent can optimize economic and environmental performance in unison without the need for explicit rule-based programming.

5. Discussion

This study shows that the suggested physics-informed MO-PPO framework based on the DT with the use of the DT data driver can deliver stable and statistically significant changes in manufacturing performance in the dimensions of reliability, sustainability, and adaptability. The framework demonstrated better schedule compliance of 96.8%, general OEE of 84.7%, specific energy usage of 2.38 kWh/kg, material waste rate of 6.8 percent, and a Pareto hyper-volume of 0.84, which is significantly better than all benchmark techniques with $p < 0.001$. These profits were maintained with extreme perturbations, such as machine failures, energy–carbon shocks, and composite stress cases, which means that the proposed approach not only optimizes on nominal conditions, but is also robust to the actual operational uncertainty.

Some of the findings, however surprising, have serious theoretical implications. Against the current belief that sustainability goals necessarily affect the operational reliability, our findings reveal that there is a strong synergetic correlation between waste minimization and the OEE improvement (Pearson r = −0.73, $p < 0.001$). Using regression analysis, we can infer that 34.1% of OEE improvements can be directly linked to waste-reducing approaches with mediation by an extended tool life (+23%), a reduced vibration RMS (−34%), a higher first-pass yield (+0.5 pp), and a reduced number of unplanned failures (−31%). These results undermine the conventional Pareto-conflict paradigm of the context of static multi-objective optimization and suggest that sustainability interventions can constructively contribute to reliability, and not to undermine it, when explicitly represented in models, and when the physics of degradation and energy dynamics are explicitly included in them.

The given framework is superior to the former approaches because it allows for optimizing multiple objectives in real time and stability. NSGA-II creates competitive schedules; however, it takes 14–15 min, which makes it inappropriate in minute-level replanning [31,57,64]. The omissions of sustainability modeling in single-objective DQN overfit throughput, and is 22 percent more energy-consuming during energy–carbon volatility [23,27,29,35]. A3C enhances OEE but generates more waste by about 12%, since sustainability constraints have not been considered [26,43,47]. MO-DQN has lower Pareto performance (0.71 compared to 0.84) and requires 2.4 times the number of training steps because of ineffective exploration [24,57,65]. In contrast, MO-PPO facilitates higher convergence, higher Pareto diversity, and stable trade-offs using Pareto-conditioned learning and replay, and is hence consistent with recent developments in DT and RL [7,52,53,58,64].

The three interdependent mechanisms can be the reason behind the observed improvements in performance. To begin with, the state encoder of the Graph Attention Network (GAT) provides the topology of the shop floor and its long-range dependencies; the ablation experiments, which substituted the GAT with a multilayer perceptron, led to a

reduction in the performance by 9 percent, thus confirming the significance of relational awareness. Second, by using preference conditioning over twelve ω vectors, Pareto diversity enforcing allows collapse to single-objective behavior to be avoided, whereas with MO-DQN, a fraction of 84 percent of the gainable Pareto hypervolume is covered. Third, risk-aware exploration in a high-fidelity physics-informed DT allows the agent to undergo more than 400 simulated failures in training, thus learning risk-aware policies that predict remaining useful life within $\pm$2.1 h; real-world deployment had no safety incidents. This realistic-to-safe combination is vital in the transfer of simulation to reality.

The framework is relatively expensive ($\approx$12 GPU hours) to perform the necessary calculations, depends on six months of historical data to calibrate DT, and has been tested on a single cell of production, with multi-cell coordination identified as an important direction for future work. Legacy machines needed quantization to accommodate continuous control actions, and initial deployment was constrained by operator acceptance. Despite these limitations, the proposed method demonstrates robust empirical performance, including low inference latency (47 ms), low simulation-to-reality error (2.8%), and seamless integration with the MES. Overall, work provides a quantitative and system-level contribution by demonstrating how sustainability and reliability can be jointly optimized within an adaptive, Industry 5.0-oriented manufacturing planning framework.

The proposed framework is tested in a job-shop manufacturing environment, and the state representation is accordingly tailored to discrete scheduling decisions and machine degradation dynamics. Nevertheless, the underlying DT-reinforcement learning structure and the limited multi-objective model are not domain-specific. Extending the framework to other manufacturing paradigms would primarily imply redefining the action and state spaces. In continuous-flow or process-oriented industries, such as chemical manufacturing, job-level variables can be replaced by process-level indicators, including throughput rates, residence time, energy intensity, and process stability. This flexibility enables broad applicability across different manufacturing contexts while preserving the adaptive decision-making core of the proposed framework.

6. Conclusions

This study presents a machine learning-based adaptive planning system that can illustrate how both sustainability and operational reliability can be optimized in a mutually compatible and effective way in real-life manufacturing systems. Unifying a physics-based DT and a Pareto-conditioned MORL allows the proposed structure to enable real-time, closed-loop decision making under uncertainty whilst being compatible with industry execution systems. Extensive simulation testing and a long-term industrial pilot project can confirm that the platform provides robust, reliable, and understandable performance gains in a wide range of disruption conditions, and does not violate operational safety and constraint fulfillment. The positive performance effects and the results give empirical support that the environmental goals, e.g., energy efficiency and waste reduction, can enhance, instead of weaken, reliability in the case of explicitly modeled degradation dynamics and system interactions. The findings demonstrate the significance of state holistic representation, safe exploration, and policy diversity to implement learning-based control in cyber–physical manufacturing systems. Altogether, this piece of work outlines a viable roadmap to adaptive, resilient, and low-impact manufacturing processes and provides a basis for future research on scalable, human-centered, and federated intelligent planning architectures in the framework of Industry 5.0.

Author Contributions: Conceptualization, M.L. and C.-M.Y.; methodology, C.-M.Y. and Y.-W.K.; software, Y.-W.K.; validation, Y.-W.K.; formal analysis, C.-M.Y.; data curation, W.L.; writing—original draft preparation, M.L., C.-M.Y. and W.L.; writing—review and editing, M.L., C.-M.Y. and Y.-W.K.;

visualization, C.-M.Y. and W.L.; funding acquisition, M.L. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Education Humanities and Social Sciences Youth Fund Project, grant number 22YJC630060, the Research Project Funded for the Construction of Guangxi's First Class Discipline Applied Economics (Digital Economy Direction), grant number 2024GSXKB04, and the 75th Batch of General Projects of China Postdoctoral Science Foundation, grant number 2024M750476.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lee, J.; Bagheri, B.; Kao, H.-A. A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [CrossRef]
2. Kusiak, A. Smart manufacturing. *Int. J. Prod. Res.* **2018**, *56*, 508–517. [CrossRef]
3. Machado, C.G.; Winroth, M.P.; da Silva, E.H.D.R. Sustainable manufacturing in Industry 4.0: An emerging research agenda. *Int. J. Prod. Res.* **2020**, *58*, 1462–1484. [CrossRef]
4. Ghobakhloo, M. Industry 4.0, digitization, and opportunities for sustainability. *J. Clean. Prod.* **2020**, *252*, 119869. [CrossRef]
5. Barua, D.A.; Sami, S.A.; Barua, L. Leveraging artificial intelligence for smart production management in Industry 4.0. *Sci. Rep.* **2025**, *15*, 41559. [CrossRef]
6. *ISO 50001:2018*; Energy Management Systems—Requirements with Guidance for Use. International Organization for Standardization: Geneva, Switzerland, 2018.
7. Lu, Y.; Liu, C.; Wang, K.I.; Huang, H.; Xu, X. Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues. *Robot. Comput.-Integr. Manuf.* **2020**, *61*, 101837. [CrossRef]
8. Zhong, R.Y.; Xu, X.; Klotz, E.; Newman, S.T. Intelligent manufacturing in the context of Industry 4.0: A review. *Engineering* **2017**, *3*, 616–630. [CrossRef]
9. Bokrantz, J.; Skoogh, A.; Berlin, C.; Wuest, T.; Stahre, J. Smart maintenance: A research agenda for industrial maintenance management. *Int. J. Prod. Econ.* **2020**, *224*, 107547. [CrossRef]
10. Monostori, L.; Kádár, B.; Bauernhansl, T.; Kondoh, S.; Kumara, S.; Reinhart, G.; Sauer, O.; Schuh, G.; Sihn, W.; Ueda, K. Cyber-physical systems in manufacturing. *CIRP Ann.* **2016**, *65*, 621–641. [CrossRef]
11. Stock, T.; Obenaus, M.; Kunz, S.; Kohl, H. Industry 4.0 as enabler of sustainable development: A qualitative assessment of its ecological and social potential. *Process Saf. Environ. Prot.* **2018**, *118*, 254–267. [CrossRef]
12. Morelli, G.; Magazzino, C.; Gurrieri, A.R.; Pozzi, C.; Mele, M. Designing smart energy systems in an industry 4.0 paradigm towards sustainable environment. *Sustainability* **2022**, *14*, 3315. [CrossRef]
13. Kamble, S.S.; Gunasekaran, A.; Gawankar, S.A. Sustainable Industry 4.0 framework: A systematic literature review. *Process Saf. Environ. Prot.* **2018**, *56*, 254–271. [CrossRef]
14. Dalenogare, L.S.; Benitez, G.B.; Ayala, N.F.; Frank, A.G. The expected contribution of Industry 4.0 technologies to industrial performance. *Int. J. Prod. Econ.* **2018**, *204*, 383–394. [CrossRef]
15. de Sousa Jabbour, A.B.L.; Jabbour, C.J.C.; Godinho Filho, M.; Roubaud, D. Industry 4.0 and the circular economy: A proposed framework for sustainable operations. *Prod. Plan. Control* **2018**, *29*, 576–586. [CrossRef]
16. Okorie, O.; Salonitis, K.; Charnley, F.; Tiwari, A. Digitisation and the circular economy: A review of current research and future trends. *Energies* **2018**, *10*, 3009. [CrossRef]
17. Thoben, K.D.; Wiesner, S.; Wuest, T. Industrie 4.0 and smart manufacturing—A review of research issues and application examples. *Int. J. Autom. Technol.* **2017**, *11*, 4–16. [CrossRef]
18. Lee, C.G.; Park, S.C. Survey on the virtual commissioning of manufacturing systems. *J. Comput. Des. Eng.* **2014**, *1*, 213–222. [CrossRef]
19. Ivanov, D.; Dolgui, A.; Sokolov, B. The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *Int. J. Prod. Res.* **2019**, *57*, 829–846. [CrossRef]
20. Mittal, S.; Khan, M.A.; Romero, D.; Wuest, T. A critical review of smart manufacturing and Industry 4.0 maturity models: Implications for small and medium-sized enterprises. *J. Manuf. Syst.* **2018**, *49*, 194–214. [CrossRef]
21. Sony, M.; Naik, S. Ten lessons for managers implementing Industry 4.0. *IEEE Eng. Manag. Rev.* **2019**, *47*, 45–52. [CrossRef]
22. Soori, M.; Ghaleh Jough, F.K.; Dastres, R.; Arezoo, B. AI-based decision support systems in Industry 4.0: A review. *J. Econ. Technol.* **2024**, *4*, 101253. [CrossRef]

23. Esteso, A.; Peidro, D.; Mula, J.; Díaz-Madroñero, M. Reinforcement learning applied to production planning and control. *Int. J. Prod. Res.* **2023**, *61*, 5772–5789. [CrossRef]
24. Chen, B.; Zha, J.; Cai, Z.; Wu, M. Predictive modelling of surface roughness in precision grinding based on a hybrid algorithm. *CIRP J. Manuf. Sci. Technol.* **2025**, *59*, 1–17. [CrossRef]
25. Wu, M.; Arshad, M.H.; Saxena, K.K.; Qian, J.; Reynaerts, D. Profile prediction in ECM using machine learning. *Procedia CIRP* **2022**, *113*, 410–416. [CrossRef]
26. Modrák, V.; Sudhakarapandian, R.; Balamurugan, A.; Soltysova, Z. A review on reinforcement learning in production scheduling: An inferential perspective. *Algorithms* **2024**, *17*, 343. [CrossRef]
27. Achamrah, F.E.; Attajer, A. Multi-objective reinforcement learning-based framework for solving selective maintenance problems in reconfigurable cyber-physical manufacturing systems. *Int. J. Prod. Res.* **2024**, *62*, 3460–3482. [CrossRef]
28. Johnson, D.; Chen, G.; Lu, Y. Multi-agent reinforcement learning for real-time dynamic production scheduling in a robot assembly cell. *IEEE Robot. Autom. Lett.* **2022**, *7*, 7684–7691. [CrossRef]
29. del Real Torres, A.; Andreiana, D.S.; Roldan, A.O.; Bustos, A.H.; Galicia, L.E.A. A review of deep reinforcement learning approaches for smart manufacturing in industry 4.0 and 5.0 framework. *Appl. Sci.* **2022**, *12*, 12377. [CrossRef]
30. Vespoli, S.; Mattera, G.; Marchesano, M.G.; Nele, L.; Guizzi, G. Adaptive manufacturing control with deep reinforcement learning for dynamic WIP management in industry 4.0. *Comput. Ind. Eng.* **2025**, *202*, 110966. [CrossRef]
31. Chang, J.; Yu, D.; Hu, Y.; He, W.; Yu, H. Deep reinforcement learning for dynamic flexible job shop scheduling with random job arrival. *Processes* **2022**, *10*, 760. [CrossRef]
32. Wu, M.; Yao, Z.; Verbeke, M.; Karsmakers, P.; Gorissen, B.; Reynaerts, D. Data-driven models with physical interpretability for real-time cavity profile prediction in electrochemical machining processes. *Eng. Appl. Artif. Intell.* **2025**, *160*, 111807. [CrossRef]
33. Moosavi, S.; Farajzadeh-Zanjani, M.; Razavi-Far, R.; Palade, V.; Saif, M. Explainable AI in manufacturing and industrial cyber–physical systems: A survey. *Electronics* **2024**, *13*, 3497. [CrossRef]
34. Lienenlüke, L.; Storms, S.; Brecher, C. Predicting Path Inaccuracies in Robot-based Machining Operations Using Inverse Kinematics. *IFAC-PapersOnLine* **2019**, *52*, 1785–1790. [CrossRef]
35. Del Gallo, M.; Mazzuto, G.; Ciarapica, F.E.; Bevilacqua, M. Artificial intelligence to solve production scheduling problems in real industrial settings: Systematic literature review. *Electronics* **2023**, *12*, 4732. [CrossRef]
36. Zhang, Y.; Huang, G.Q.; Sun, S.; Yang, T. Multi-agent based real-time production scheduling method for radio frequency identification enabled ubiquitous shopfloor environment. *Comput. Ind. Eng.* **2014**, *76*, 89–97. [CrossRef]
37. Green, K.W.; Inman, R.A.; Sower, V.E.; Zelbst, P.J. Impact of JIT, TQM and green supply chain practices on environmental sustainability. *J. Manuf. Technol. Manag.* **2018**, *30*, 26–47. [CrossRef]
38. Li, B.H.; Hou, B.C.; Yu, W.T.; Lu, X.B.; Yang, C.W. Applications of artificial intelligence in intelligent manufacturing: A review. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 86–96. [CrossRef]
39. Kasie, F.M.; Bright, G.; Walker, A. Decision support systems in manufacturing: A survey and future trends. *J. Model. Manag.* **2017**, *12*, 432–454. [CrossRef]
40. Ding, C.; Qiao, F.; Wang, D.; Liu, J. Adaptive real-time scheduling for production and maintenance: Integrating RUL prediction with multi-agent deep reinforcement learning. *Reliab. Eng. Syst. Saf.* **2025**, *264*, 111394. [CrossRef]
41. Scharmer, V.M.; Vernim, S.; Horsthofer-Rauch, J.; Jordan, P.; Maier, M.; Paul, M.; Schneider, D.; Woerle, M.; Schulz, J.; Zaeh, M.F. Sustainable manufacturing: A review and framework derivation. *Sustainability* **2024**, *16*, 119. [CrossRef]
42. Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; Francisco, R.D.; Basto, J.P.; Alcalá, S.G. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024. [CrossRef]
43. Selcuk, S. Predictive maintenance, its implementation and latest trends. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2017**, *231*, 1670–1679. [CrossRef]
44. Wu, M.; Shukla, S.; Vrancken, B.; Verbeke, M.; Karsmakers, P. Data-driven approach to identify acoustic emission source motion and positioning effects in laser powder bed fusion with frequency analysis. *Procedia CIRP* **2025**, *133*, 531–536. [CrossRef]
45. Beier, G.; Ullrich, A.; Niehoff, S.; Reißig, M. Industry 4.0: The future of sustainable manufacturing? *J. Manuf. Technol. Manag.* **2020**, *31*, 975–993. [CrossRef]
46. Lee, J.; Davari, H.; Singh, J.; Pandhare, V. Industrial artificial intelligence for Industry 4.0-based manufacturing systems. *Manuf. Lett.* **2018**, *18*, 20–23. [CrossRef]
47. Karim, M.R. Optimizing maintenance strategies in smart manufacturing: A systematic review of lean practices, total productive maintenance (TPM), and digital reliability. *Rev. Appl. Sci. Technol.* **2025**, *4*, 176–206. [CrossRef]
48. He, Y.; Han, X.; Gu, C.; Chen, Z. Cost-oriented predictive maintenance based on mission reliability state for cyber manufacturing systems. *Adv. Mech. Eng.* **2018**, *10*, 1687814017751467. [CrossRef]
49. Abadi, A.; Abadi, C.; Abadi, M. Artificial intelligence and digital twins for sustainable production systems. *Sens. Transd.* **2025**, *270*, 1–10.

50. Prasara-A, J.; Gheewala, S.H. An assessment of social sustainability of sugarcane and cassava cultivation in Thailand. *Sustain. Prod. Consum.* **2021**, *27*, 372–382. [CrossRef]
51. Mourtzis, D.; Angelopoulos, J.; Panopoulos, N. Robust engineering for the design of resilient manufacturing systems. *Appl. Sci.* **2021**, *11*, 3067. [CrossRef]
52. Leng, J.; Ruan, X.; Jiang, P.; Xu, K.; Liu, Q.; Zhou, X.; Liu, C. Blockchain-empowered sustainable manufacturing and product lifecycle management in Industry 4.0: A survey. *Renew. Sustain. Energy Rev.* **2020**, *132*, 110112. [CrossRef]
53. Vrignat, P.; Kratz, F.; Avila, M. Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108140. [CrossRef]
54. Ghadge, A.; Er Kara, M.; Moradlou, H.; Goswami, M. The impact of Industry 4.0 implementation on supply chains. *J. Manuf. Technol. Manag.* **2020**, *31*, 669–686. [CrossRef]
55. Lee, J.; Kao, H.-A.; Yang, S. Service innovation and smart analytics for Industry 4.0 and big data environment. *Procedia CIRP* **2014**, *16*, 3–8. [CrossRef]
56. Zhang, L.; Yan, Y.; Hu, Y.; Ren, W. Reinforcement learning and digital twin-based real-time scheduling method in intelligent manufacturing systems. *IFAC-PapersOnLine* **2022**, *55*, 359–364. [CrossRef]
57. Xia, K.; Sacco, C.; Kirkpatrick, M.; Saidy, C.; Nguyen, L.; Kircaliali, A.; Harik, R. A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. *J. Manuf. Syst.* **2020**, *58*, 210–230. [CrossRef]
58. Khdoudi, A.; Masrour, T.; El Hassani, I.; El Mazgualdi, C. A deep-reinforcement-learning-based digital twin for manufacturing process optimization. *Systems* **2024**, *12*, 38. [CrossRef]
59. Pavlenko, P.; Yu, B. Digital twin and reinforcement learning-based additive manufacturing optimization. In Proceedings of the 4th International Conference on Electronic Information Engineering and Computer Science, Yanji, China, 27–29 September 2025; p. 13574. [CrossRef]
60. Paranjape, A.; Quader, N.; Uhlmann, L.; Berkels, B.; Wolfschläger, D.; Schmitt, R.H.; Bergs, T. Reinforcement learning agent for multi-objective online process parameter optimization of manufacturing processes. *Appl. Sci.* **2025**, *15*, 7279. [CrossRef]
61. Zhang, L.; Qi, Z.; Shi, Y. Multi-objective reinforcement learning–concept, approaches and applications. *Procedia Comput. Sci.* **2023**, *221*, 526–532. [CrossRef]
62. Tao, F.; Qi, Q.; Wang, L.; Nee, A.Y.C. Digital twins and cyber–physical systems toward smart manufacturing and Industry 4.0: Correlation and comparison. *Engineering* **2019**, *5*, 653–661. [CrossRef]
63. Tao, F.; Cheng, J.; Qi, Q.; Zhang, M.; Zhang, H.; Sui, F. Digital twin-driven product design, manufacturing and service with big data. *Int. J. Adv. Manuf. Technol.* **2018**, *94*, 3563–3576. [CrossRef]
64. Gao, P.; Li, X.; Yan, X.; Li, H.; Zhan, M. Digital twin-driven intelligent spinning technique for curved surface parts. *J. Ind. Inf. Integr.* **2025**, *45*, 100848. [CrossRef]
65. Leng, B.; Gao, S.; Xia, T.; Pan, E.; Seidelmann, J.; Wang, H.; Xi, L. Digital twin monitoring and simulation integrated platform for reconfigurable manufacturing systems. *Adv. Eng. Inform.* **2023**, *58*, 102141. [CrossRef]
66. Liu, M.; Fang, S.; Dong, H.; Xu, C. Review of digital twin about concepts, technologies, and industrial applications. *J. Manuf. Syst.* **2021**, *58*, 346–361. [CrossRef]
67. Wu, Z.; Li, J. A framework of dynamic data driven digital twin for complex engineering products: The example of aircraft engine health management. *Procedia Manuf.* **2021**, *55*, 139–146. [CrossRef]
68. Zhuang, C.; Miao, T.; Liu, J.; Xiong, H. The connotation of digital twin, and the construction and application method of shop-floor digital twin. *Robot. Comput.-Integr. Manuf.* **2021**, *68*, 102075. [CrossRef]
69. Gao, Y.; Lou, S.; Zheng, H.; Tan, J. A data-driven method of selective disassembly planning at end-of-life under uncertainty. *J. Intell. Manuf.* **2023**, *34*, 565–585. [CrossRef]
70. Sresakoolchai, J.; Kaewunruen, S. Railway infrastructure maintenance efficiency improvement using deep reinforcement learning integrated with digital twin. *Sci. Rep.* **2019**, *11*, 3803. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.