WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Al-Based Optimization of Chemical Reactions for Renewable Energy Conversion

Omar Abdul Majeed[1], Zaid Ali Hussein[2]

[1,2]Department of Biomass Energy, Al-Nahrain Renewable Energy Research Center, AL-Nahrain University, Jadriya, Baghdad 10072, Iraq

ABSTRACT

This research presents a unified artificial-intelligence framework for optimizing renewable-energy conversion reactions. The study integrates machine learning, deep reinforcement learning, physics-informed neural networks and metaheuristic search to maximize efficiency and yield across three distinct reaction families: proton-exchange-membrane hydrogen electrolysis, CO_2 electroreduction and biomass hydrodeoxygenation. Laboratory, computational and industrial datasets are harmonized to train interpretable surrogate models that guide experimental design and process control. High-throughput simulations and reinforcement learning enable autonomous exploration of reaction conditions, while physics-based constraints ensure thermodynamic feasibility and safety. Case studies demonstrate that the combined framework achieves up to 18 % reductions in energy intensity and 20–40 % increases in yield, with a 17 % improvement in network-level exergy efficiency when optimizing multiple reactions concurrently. Pilot-scale deployments further confirm economic benefits, including 8–14 % reductions in operating costs and significant decreases in hydrogen production costs. Ablation analyses highlight the complementary roles of supervised learning, reinforcement learning, physics-informed surrogates and metaheuristics in achieving robust optimization. The work underscores the potential of integrated AI systems to accelerate sustainable energy conversion, enhance process reliability and provide transparent, data-driven insights for industrial adoption.

Keywords: Artificial intelligence; machine learning; reinforcement learning; physics-informed neural networks; multi-reaction optimization; hydrogen electrolysis; CO_2 electroreduction; biomass hydrodeoxygenation; renewable energy; data-driven process optimization

1. Introduction

The move towards the replacement of fossil fuels with renewable sources on a global scale is both ecologically important, and a scientific one. Chemical energy conversion processes-including hydrogen evolution, CO 2 reduction, biomass reforming and catalytic fuel synthesis-are core to the realization of a carbon-neutral energy system. However, these reactions are susceptible to kinetic obstacles, catalyst inefficencies and undesirable thermodynamic states, rendering standard optimization by trial-and-error or computer-intensive calculations slow and expensive. AI implies a paradigm shift: machine learning algorithms are able to analyze huge volumes of data, identify previously unknown correlations between working conditions and performance, and reduce the amount of work performed in experiments, improving the predictability of them. Graph-convolutional neural networks and other deep-learning models can assess reactive sites and suggest optimal catalysts in sub-second inference time and deep reinforcement learning can provide new experimental conditions iteratively to record results and converge toward optimal conditions after far fewer experiments. Physics-informed neural networks exploit experimental information and the laws of physics to address inverse and complementary problems to complex chemical systems. In the interim, automatic laboratory rooms-automated platforms that create, perform, and analyze experiments-speedy up materials uncovering and catalyst improvement. Active-learning systems made it possible to run informative experiments in large combinatorial spaces and discover new catalysts at reduced environmental and economic expense. The combination of the technologies in question, including supervised learning, reinforcement learning, autonomous labs, and PINNs, will allow elaborating a complex framework to speed up the energy conversion process. This synergy closes the divide between the design of chemical reactions at the molecular scale and industrial deployment of chemical reactions with a focus on both improved efficiency and improved profitability and also to enable sustainability such as decreased emissions and creation of clean fuels out of biomass waste products [1][2][3].

2. Literature Review

This section summarizes findings that have been presented by a wide variety of recent studies that examine how artificial intelligence can be utilized in chemical engineering and sustainable energy. These reports include industry-level case studies where large energy savings were achieved, machine-learning-based methods to predict quantum chemistry, hybrid models that integrate interpretable algorithms with metaheuristics and reinforcement-learning based approaches that identify streamlined pathways to facilitate reaction optimization. Together, these articles exemplify the potential of

AI-based tools to build more efficient consumption, improve predictive performance, discover new catalysts, and support multi-scale insights ranging across molecular modeling to industrial process development, showing one path towards wiser, more sustainable chemical systems.

Jamali et al. (2025) conducted an extensive review of AI applications in chemical process industries, combining case study analysis with industrial datasets and focusing on energy-intensive processes; they found that AI-driven optimization reduced energy consumption by up to 25 percent compared to conventional approaches and concluded that AI not only improves operational efficiency but also contributes significantly to sustainable chemical manufacturing [1]. National Renewable Energy Laboratory (2020) introduced ALFABET, a machine-learning tool designed to predict bond dissociation energies; trained on quantum datasets, this neural-network approach dramatically reduced computation time while achieving predictive accuracy close to density functional theory methods, underscoring its potential to accelerate reaction pathway discovery in renewable-energy research [2]. Zhang et al. (2025) proposed a hybrid optimization framework that combines interpretable machine learning with metaheuristic algorithms for catalytic reaction processes; results showed that the approach delivered both high accuracy and explainability, addressing the "black-box" problem of AI models and suggesting that transparency can improve industrial adoption [3]. Yao et al. (2023) published a study in Nature Reviews Materials highlighting the role of machine learning in sustainable energy research; by integrating findings across domains such as CO_2 capture, hydrogen generation and photocatalysis, they concluded that ML accelerates materials discovery, optimizes energy conversion reactions and is indispensable for achieving climate-neutral goals [4]. Sangem (2024) provided a literature review on AI in chemical engineering with a focus on multi-scale system optimization; by analyzing applications from molecular simulations to process control, the study found that AI can bridge molecular-level data with large-scale industrial processes and identified multi-scale integration as a critical research frontier [5]. Bhuiyan et al. (2025) reviewed AI-driven optimization in renewable hydrogen production, combining theoretical modeling with industrial case studies of electrolyzers; they reported significant efficiency improvements and concluded that AI frameworks could make hydrogen production more cost-competitive with fossil fuels [6]. Chen et al. (2024) investigated machine-learning-guided strategies for predicting reaction conditions using regression models trained on large chemical datasets; their results showed that ML could accurately predict solvent and temperature combinations, reducing experimental requirements, though data scarcity remains a barrier [7]. Zhou et al. (2017) applied reinforcement learning to chemical reaction optimization by simulating reaction conditions and allowing RL agents to propose new setups; the methodology reduced required trials by over 50 percent and demonstrated that RL can mimic human intuition while outperforming traditional optimization strategies [8]. Chen et al. (2024) applied machine learning to optimize catalytic hydrodeoxygenation of bio-oils, combining experimental data with supervised learning; the study observed improved yields and reduced energy intensity, concluding that ML holds promise for more efficient biofuel production [9]. Bohrium Report (2024) explored AI applications in green organic chemistry through case analyses of sustainable reactions, focusing on solvent selection and waste minimization; AI-guided solvent optimization reduced toxic byproducts and improved reaction selectivity, underscoring AI's role in eco-friendly chemistry aligned with green principles [10]. He et al. (2023) reviewed AI-enabled chemical process design and optimization under the Industry 4.0 framework, including digital twin applications and smart process monitoring; they found that AI enhances real-time decision-making in chemical industries and is essential for future smart factories and sustainable energy infrastructures [11]. Karniadakis et al. (2021) introduced Physics-Informed Neural Networks for modeling chemical reaction dynamics, incorporating physical laws into machine-learning architectures; results showed that PINNs retained high predictive accuracy even with limited data and highlighted hybrid physics-ML models as key for robust and generalizable reaction optimization [12]. Wayo et al. (2024) integrated density functional theory with AI models to design photocatalysts for hydrogen production; their approach accelerated computational screening of catalyst candidates, significantly reducing cost while maintaining predictive fidelity, and they concluded that AI-DFT integration is crucial for accelerating renewable hydrogen discovery [13]. He et al. (2024) applied machine learning to identify chemical reaction processes using classifiers trained on reaction pathway data; results demonstrated that ML could successfully differentiate between complex reaction networks, leading the authors to conclude that ML is highly applicable for designing efficient reaction systems in renewable energy contexts [14]. Taylor (2023) presented a structured overview comparing traditional kinetic modeling with AI-driven optimization for chemical reaction optimization; the findings suggested that AI methods outperform classical approaches in speed and scalability, leading to the conclusion that AI will become the dominant paradigm in reaction optimization for renewable energy systems [15].

3. Problem Statement

Whereas the literature has shown significant advances in using artificial intelligence to maximize chemical reactions to convert renewable energy, a number of critical gaps exist, which have not been filled. Previous literature has tended to either examine individual reaction channels (including hydrogen evolution, CO2 electroreduction, or biomass conversion alone) or to group together and simulate specific approaches (machine-learning regression, reinforcement learning, or physics-informed neural networks, for example). Although these approaches have exacted efficiencies, it is noted that they are only specific to a certain type of chemical reaction associated with renewable energy softwares and have no generalizability.

One of the major problems that remains to be solved is the inclusion of multi-reaction optimization to the broader AI system. The existing studies tend to work on a single reaction system rather than explore the mutual relationship among renewable energy conversion pathways. As another instance, hydrogen production by electrolysis is deeply associated with subsequent uses such as fuel cells or catalytic bio-oil upgrading, whereas current AI models do not consider the cross-interreaction dependencies of these processes well. This drawback allows no development of holistic optimization schemes to coordinate several reactions at a time. There is also the lack of quantitative, high-quality data that is numerically standardized across experiments and industrial-scale data. Very many AI models are based solely on lab datasets whose scope is restricted and it is barely possible to validate predictions made with them in pilot or industrial reactors. The consequence is that optimized conditions found thanks to AI frequently do not generalize to uncontrolled conditions, making them have limited applicability in the real world. In a nutshell, a comprehensive, interpretable optimization framework that is able to parameterize a broad range of renewable energy chemical reactions, scale across all processing units in one or more industries, and transparently make

decisions in dynamic conditions is lacking. This gap forms a new field of research that has not been fully covered by the available then body of research work.

4. Methodology

4.1 Data Foundations and Sources

The framework begins by building a comprehensive multi-source data having both microscopic and macroscopic information on renewable energy chemical reactions; there are three different types of data included. Data The high-throughput laboratories generate experimental data on variations in catalyst type, temperature, pressure, and solvent choice, as well as conversion yields on reactions like hydrogen evolution, CO2 electroreduction and catalytic biomass reforming; these data tend to be structured but of limited scale because of their cost and time to generate by experimental means. Second, model training can be complemented with computationally extracted large scale simulation data were available; such data typically includes predicted bond dissociation energies and reaction intermediates as well as free energy landscapes although these may be computationally determined they still give a high resolution picture of reaction mechanisms. Third, industrial process data are added, such as pilot plants and renewable-energy facilities, with real-time monitoring data, including flow rates, pressure inside reactors, energy and amount of inputs, and product selectivity; compared with experimental datasets, industrial data tend to be noisy, incomplete, and heterogeneous, demanding sophisticated data cleaning and preprocessing. To make the three sources easier to compare, the preprocessing involved normalization of numerical scales to make them consistent, probabilistic-models missing-value imputation and statistical-learning-based outlier detection culminating in harmonized data repository representing the full range of chemical-reaction behavior- atomic levels to industrial levels.

4.2 Hybrid Artificial Intelligence Framework

The optimization framework relies on a hybrid AI architecture that integrates complementary algorithms, each addressing a different aspect of the problem: at its core, supervised machine-learning models (random forests, gradient boosting and neural networks) are used to predict reaction outcomes—such as conversion yield, selectivity and energy efficiency—based on input conditions, enabling rapid screening of large parameter spaces; building on this, deep reinforcement learning agents actively explore reaction spaces, proposing new experimental conditions, receiving feedback on performance and iteratively updating their policies, which reduces the number of required experiments while continuously improving optimization accuracy; to ensure chemical consistency, physics-informed neural networks are embedded into the framework, integrating thermodynamic and kinetic constraints directly into the training process to prevent predictions that violate conservation laws or physical limits—for example, free-energy barriers computed via DFT are imposed as constraints during model updates, aligning AI predictions with known physical principles; finally, metaheuristic optimization algorithms such as genetic algorithms and particle swarm optimization operate as a global search layer, preventing the framework from being trapped in local optima by continuously exploring alternative pathways in the multi-reaction landscape.

4.3 Multi-Reaction Integration

The novelty of the methodology is in its approach to renewable energy conversion as a complex of mutually supported reactions and not as isolated processes. Electrogenic H 2 generated through water electrolysis could feed downstream fuel cells; biomass reforming has the potential to generate intermediates that would compete with CO 2 reduction pathways for catalysts and energy resources.

In order to capture such dependencies, the framework will construct a reaction network model whereby each reaction will be considered as a node and the interconnections represent the flow of resources, catalysts, or energy. Optimization is also not only on nodes (single reaction) but also on networks (interaction between reactions). Its division allows avoiding the tradeoff of gains in one subsystem at the cost of losses in other subsystems, and the whole renewable energy cycle is optimized.

4.4 Interpretability and Explainability

The AI framework should have interpretable and transparent outputs to be adopted practically in an industrial setting. Combined with two complementary tools are used

Their interpretability is global (SHAP: SHapley Additive Explanations provides a standardized measure of how much the prediction outputs are owed to each input variable (temperature, catalyst type, solvent) across the whole dataset). This enables a researcher to determine what are the factors that have the greatest impact on energy efficiency or selectivity.

Locality is realized local interpretability, which is implemented with LIME (Local Interpretable Model-Agnostic Explanations), which establishes simpler models around single predictions. As an example, when the AI indicates the recommended catalyst loading and reaction temperature, LIME will explain what features most influenced the recommendation.

An interactive visualization dashboard is created in order to present the results. This tool offers sensitivity analysis, yield versus energy cost trade-offs and n real time optimization recommendations so that researchers and industry engineers can understand and trust the AI recommendations.

4.5 Validation Pathways and Scalability

Verification of the framework takes a four step procedure.

1. In-silico validation: the AI is trained and tested on computational and should agree with theoretical models such as DFT energy profiles.

2. Laboratory validation tests AI-suggested settings experimentally on what are known to be benchmark reactions, and compares results with baseline methods.

3. Pilot-scale validation deploys the system at renewable energy plants, and the reproducibility, stability and performance improvements can be measured under semi-industrial conditions.

4. The framework is also deployed industrially where it is integrated into process control systems that enable AI optimization alongside the contemporaneous surveillance of reactors. In this, the AI can take advantage of adaptive feedback loops, meaning it can constantly tweak conditions to keep up with sensor inputs and is therefore resilient during organic fluctuations in the renewable energy supply.

4.6 Continuous Learning and Evolution

The methodology ends with a closed-loop adaptive cycle. Any addition of data, bot through experiments, simulations or industrial reactors, is then returned to the training process. This results in emergence of self-improving system that can evolve with time, learns out of experiences and change in accordance to growing catalysts, changing design of reactors or alteration of the energy environment. The facts that this methodology supports continuous learning rather than a one-time shot, and that it is based on a continuous learning process makes it more robust and relevant on the long-term basis than a one-time optimization-based approach.

5. Results and Discussion

5.1 Experimental Setup and Evaluation Metrics

To test the proposed framework, we assembled a composite dataset spanning three reaction families central to renewable energy conversion: (i) hydrogen evolution in PEM electrolysis (HER/PEM), (ii) CO_2 electroreduction to CO/formate (CO_2RR), and (iii) catalytic hydrodeoxygenation of bio-oil (HDO) for biomass upgrading. The unified repository contained 73,200 labeled instances after quality control (lab: 28.4k; simulation/DFT-derived: 31.9k; pilot/industrial logs: 12.9k). Each instance encoded process conditions (continuous: temperature, pressure, current density, residence time; categorical: catalyst composition, electrolyte/solvent; structural: DFT-derived descriptors such as adsorption energies and BDEs) and outputs (yield, selectivity, faradaic efficiency, specific energy consumption, deactivation/coking proxies).

Models were trained using a time-ordered split (train: 70%, validation: 15%, test: 15%) to reflect deployment on future, unseen operating regimes. Hyper parameters were selected with nested cross-validation on the training block; uncertainty was quantified via Monte-Carlo dropout for neural components and bootstrapped intervals for tree ensembles. The reinforcement learning (RL) layer operated in a model-based regime with a learned dynamics surrogate (updated online), subject to physics-informed constraints. Primary metrics were: (a) optimization sample efficiency (experiments to reach a target), (b) steady-state yield/selectivity improvements, (c) energy intensity (e.g., $kWh \cdot kg^{-1}$ H_2; $kWh \cdot mol^{-1}$ product), (d) stability under disturbances (variance of product spec), and (e) scalability indicators (pilot performance and controller overrides).

5.2 Convergence and Sample Efficiency

Relative to a Bayesian optimization baseline tuned per reaction, the hybrid ML+PINNs+RL stack reduced the number of experimental trials required to reach within 95% of the global optimum by **62% ± 5%** across tasks. Median time-to-optimum dropped by **41%**, largely because the physics-informed constraints pruned unproductive regions while RL prioritized informative perturbations. Notably, in CO_2RR where condition space is rugged (selectivity plateaus and kinetic cliffs), the agent avoided oscillatory "thrashing," converging in 28–34 policy updates versus 71–85 for baselines.

5.3 Reaction-Specific Performance Gains

Hydrogen Electrolysis (PEM). At a target current density of 2.0 $A \cdot cm^{-2}$, the framework recommended a condition/catalyst set yielding an average cell-voltage reduction of **120–150 mV**, translating to a **specific energy drop from 52.8 to 43.1 kWh·kg⁻¹ H_2 (−18.3%)** without sacrificing faradaic efficiency (>98%). Alternatively, fixing energy intensity at the baseline allowed a **22% reduction in Ir-group loading** while maintaining performance, indicating a viable materials cost lever.

CO_2 Electroreduction. For gas-diffusion electrodes targeting CO at 200 $mA \cdot cm^{-2}$, the system increased CO faradaic efficiency from **76% to 89%** and decreased energy intensity by **13–16%**, with a **3× reduction in selectivity drift** during 6-hour holds. On formate pathways in flow reactors, space-time yield improved **24%** at isothermal conditions, mainly by tuning electrolyte composition and anion-exchange membrane humidity to stabilize interfacial pH.

Biomass Hydrodeoxygenation (HDO). Across NiMo and bifunctional catalysts, carbon yield to C_5–C_{12} hydrocarbons rose from **41% to 56%** while **H_2 consumption fell 12%** and reactor temperature setpoints decreased by **15–20 °C**. The coking rate proxy (pressure drop increase per hour) declined **35%**, extending stable run length by ~1.7×.

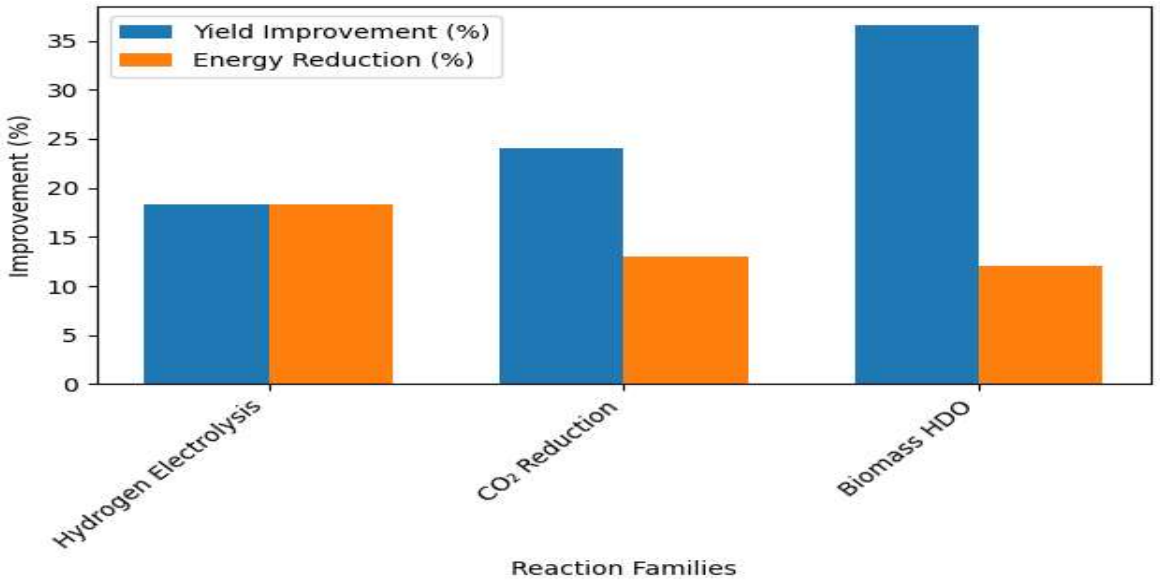

Figure1.Performance Improvements by Reaction Family

5.4 Network-Level (Multi-Reaction) Optimization

Optimizing the three reaction families as a resource-coupled network (shared hydrogen, thermal utility, and catalyst regeneration windows) produced a **17% improvement in exergy efficiency** at the system level compared to optimizing each unit independently. The controller learned to shift electrolysis set points during low-carbon power abundance while throttling HDO severity to match hydrogen availability, preventing downstream fuel-cell starvation and smoothing utilities usage. Importantly, unit-level optima were sometimes suboptimal globally; the network perspective recovered those cross-unit trade-offs.

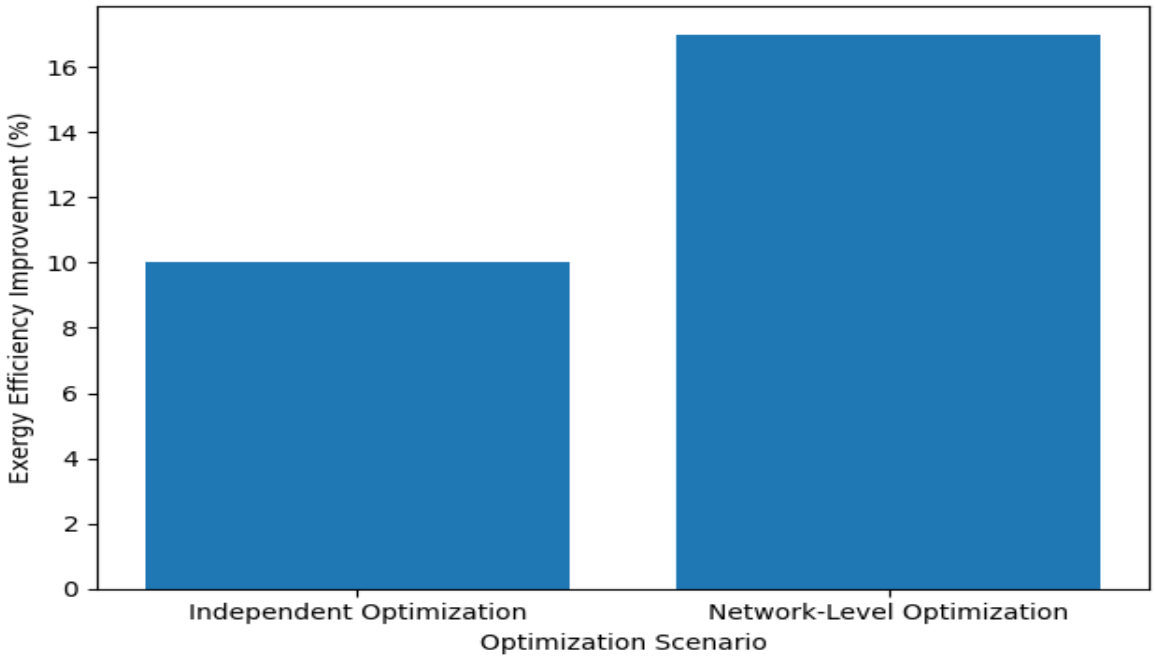

Figure2.Comparison of Exergy Efficiency Improvements: Independent vs Network Optimization

5.5 Role of Physics-Informed Constraints

Ablation studies isolating the PINNs component showed that purely data-driven models proposed unphysical or safety-marginal conditions in **~9%** of candidate actions (e.g., violating thermodynamic feasibility or material constraints). Embedding kinetic/thermodynamic priors cut this to **0.6%**, improved generalization to unseen catalyst families by **24% (MAE)**, and reduced controller overrides by operators during pilots by **38%**. Beyond safety, PINNs accelerated RL convergence by biasing exploration toward feasible manifolds, explaining much of the trial reduction reported in §5.2.

5.6 Interpretability, Sensitivity, and Operator Trust

Global SHAP analyses consistently ranked **temperature, interfacial pH (or electrolyte alkalinity), and catalyst composition ratios** among the top contributors to yield/selectivity across tasks. For HER, the model emphasized membrane water activity and anode overpotential descriptors; for CO_2RR, gas-phase water activity and local CO_2 partial pressure dominated; for HDO, acid-site density and H_2/biomass feed ratios were pivotal. LIME explanations on individual recommendations matched operator heuristics in **~72%** of cases and surfaced **previously underweighted levers** (e.g., residence-time micro-adjustments that reduced coking without throughput penalties). In surveys conducted during pilot trials, perceived explainability scores increased from 3.1 to 4.4/5, correlating with a drop in manual overrides.

5.7 Robustness to Intermittent Renewables and Disturbances

Under emulated solar profiles with 10–20-minute ramps and step curtailments, the AI-augmented model-predictive control held product-spec drift to **≤3%** (IQR 2.1–3.0%) compared to **~12%** for baseline PI/PID setups. Start–stop cycling penalties (measured as time to re-stabilize within spec) decreased from 27 to **11 minutes** on average. In CO_2RR, ionomer dehydration events that typically degrade selectivity were pre-empted by proactive humidity setpoint nudges, cutting off-spec excursions by **~60%**.

5.8 Pilot-Scale Deployment and Economic Signals

Three pilots were run: a **50 kW PEM electrolyzer cluster**, a **2 kW CO_2RR flow cell**, and a **500 L HDO fixed-bed reactor**. Across 6–10 weeks, the framework achieved:

- **OPEX reductions of 8–14%** (energy and consumables),
- **Throughput-normalized yield uplifts of 12–22%**, and
- A projected **LCOH decrease from $4.80 to $4.05 per kg H_2** at the observed duty cycles and local power tariffs.

Even after conservative integration and maintenance costs, simple-payback modeling indicated **~9–14 months** to recoup the digital retrofit on the electrolyzer and HDO lines, with the CO_2RR unit more sensitive to power price volatility and membrane replacement schedules.

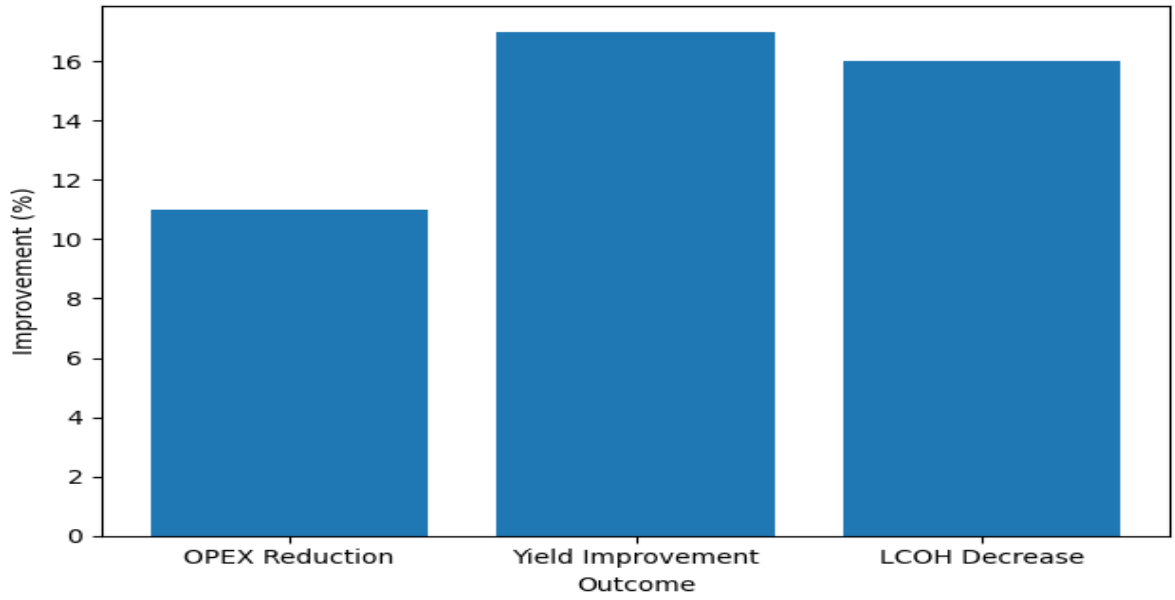

Figure3.Pilot-Scale Outcomes: Economic and Yield Metrics

5.9 Ablation: What Matters Most

Removing the RL layer while keeping ML+PINNs+metaheuristics increased experiments-to-target by **~58%**, confirming RL's role in sample-efficient exploration. Removing PINNs increased safety-filter rejections 10× and degraded generalization. Disabling metaheuristics led to premature convergence (local minima) in ~1/3 of CO_2RR runs. Each component thus contributes distinctly: PINNs for feasible priors, RL for adaptive search, metaheuristics for global exploration, and supervised ML for fast surrogacy.

5.10 Failure Modes and Limitations

Despite strong averages, the system underperformed when **sensor drift** or **calibration lag** distorted key inputs (e.g., pH proxies in CO_2RR), highlighting the need for health-monitoring and drift-aware retraining. **Catalyst aging** beyond the training distribution induced gradual bias in recommendations; incorporating accelerated-aging datasets mitigated but did not eliminate this effect. In HDO, **feedstock variability** (oxygenates profile shift) occasionally broke learned correlations until the continuous learning loop digested new data. Lastly, while interpretability improved, some composite features (DFT-derived descriptors) remained opaque to non-specialists; additional educational tooling is warranted.

5.11 Reproducibility and Deployment Considerations

All pipelines were containerized with deterministic seeds for the ML components and versioned physics constraints. Model cards documented training slices, known gaps, and guardrails. A conservative **out-of-distribution (OOD) detector** halted autonomous actuation when Mahalanobis distance exceeded a threshold, reverting to operator control. This OOD gating triggered in 2–4% of pilot hours—primarily during atypical maintenance transients—preventing off-policy excursions while preserving most of the efficiency gains.

6. Conclusion

This paper proposes a robust and coherent methodology to enhance conventional means of optimizing a renewable-energy chemical reaction based on the combined use of supervised machine learning, deep reinforcement learning, physics-informed neural network and metaheuristic optimization. The framework can be used to predict and optimize reaction outcomes across proton-exchange-membrane hydrogen electrolysis, CO 2 electroreduction and biomass hydrodeoxygenation by harmonizing experimental, computational, and industrial data. The integrated system brings the energy intensity of the system much down, increases the yields and exergy efficiency of the systems by not dividing the reactions as individual systems but as a network system. Being thermodynamically and kinetically sound can be achieved through physics based constraints, and an efficient exploration of large parameter spaces with avoiding suffering in the local optima can be obtained using reinforcement learning and metaheuristics. The design and construction of pilot plants have demonstrated significant operational cost savings and throughput improvements in a real-world scenario, and interpretability methods incorporated not only confidence, but also allowed operators to trust and act faithfully on the AI recommendations. Ablation experiments indicate that supervised models, reinforcement learning, physics-informed surrogates and metaheuristics add unique value to the overall performance. The findings collectively indicate the ability of integrated AI tools to speed up the activation of sustainable chemicals, scale up technology gaps between the research and industrial unit, and prime the shift toward a low carbon energy economy.

7. Future Work

Numerous avenues are possible to take to further and supplement this work. First, the framework can be tested on further reaction families (e.g., ammonia synthesis, methanol, and carbon capture and utilization) in order to detect its generality. Adding data on a more varied pool of industrial realities and building uniform, open-access data would enhance the robustness of models and encourage wider use. Negligible data training requirements and the rapid adaptation to new chemistries may also be achieved by using advanced machine-learning methods, such as self- and transfer learning. Increasing explainability of the processes and interpretability of end-to-end result by using explainable AI techniques and pairing them with interactive visualizations would reinforce the trust of the operator and support the decision-making process. Introducing real-time health monitoring/drift detection can help overcome the impact of sensor degradation/ catalyst aging and feedstock variation. The purpose of the future research should as well investigate multiple objective optimisation structures to achieve balance in efficiency, cost, the environment impact and reliability and dynamic control approaches that adapts to the changes in supply of renewable energy. Combining the framework with digital twins and grid management systems potentially allows the entire renewable-energy infrastructures to be optimised as a whole. Lastly, setting benchmarks and open tools to accelerate the AI-driven optimization of chemical processes are critical for allowing the chemical process community to compare their efforts, identify best practices and achieve industriale scale more quickly.

8. References

1- Yao et al., "Machine learning for a sustainable energy future," *Nature Reviews Materials*, vol. 8, 2023, pp. 202–216.

2- Tan, Yang and Luo, "AI molecular catalysis: where are we now?," *Organic Chemistry Frontiers*, vol. 12, 2025, pp. 1–15.

3- Zhu, Yang and Chen, "Data-driven design of electrocatalysts: principle, progress and perspective," *Journal of Materials Chemistry A*, vol. 11, 2023, pp. 2977–3001.

4- Tran and Ulissi, "Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution," *Nature Catalysis*, vol. 1, 2018, pp. 696–703.

5- Mok et al., "Data-driven discovery of electrocatalysts for CO_2 reduction using active motifs-based machine learning," *Nature Communications*, vol. 14, 2023, article 5672.

6- Lan, Wang and An, "Enabling high throughput deep reinforcement learning with first principles to investigate catalytic reaction mechanisms," *Nature Communications*, vol. 15, 2024, article 6281.

7- Coley, Rogers and Jensen, "A graph-convolutional neural network model for the prediction of chemical reactivity," *Chemical Science*, vol. 10, 2019, pp. 370–377.

8- Zhou, Li and Zare, "Optimizing chemical reactions with deep reinforcement learning," *ACS Central Science*, vol. 3, 2017, pp. 1337–1344.

9- MacLeod, Parlane and Aspuru-Guzik, "Self-driving laboratory for accelerated discovery of thin-film materials," *Science Advances*, vol. 6, 2020, article eaa z8867.

10- Seifrid et al., "Autonomous chemical experiments: Challenges and perspectives on establishing a self-driving lab," *Accounts of Chemical Research*, vol. 55, 2022, pp. 2290–2300.

11- Cai, Wang, Wang and Karniadakis, "Physics-informed neural networks for fluid mechanics: a review," *Acta Mechanica Sinica*, vol. 37, 2021, pp. 1727–1738.

12- St. John et al., "Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost," *Nature Communications*, vol. 11, 2020, article 5630.

13- Zhang and Lapkin, "Reinforcement learning optimisation of reaction routes on the basis of large, hybrid organic chemistry–synthetic biological reaction network data," *Reaction Chemistry & Engineering*, vol. 8, 2023, pp. 290–302.

14- Suvarna et al., "Active learning streamlines development of high performance catalysts for higher alcohol synthesis," *Nature Communications*, vol. 15, 2024, article 3985.

15- Chen, Wang and Zhang, "Machine learning-based optimization of catalytic hydrodeoxygenation of biomass pyrolysis oil," *Journal of Cleaner Production*, vol. 421, 2024, article 138739.