Chemical and Process Engineering: New Frontiers, 2026, 47(2), e139
DOI: 10.24425/cpe.2026.158137

CRITICAL REVIEW

AI/ML in Chemical Engineering: From molecular design to industrial process optimization

Fatemeh Valizadeh Hajidehi*, Azam Mina

Department of Chemical Engineering, School of Allied Medicine, Tehran University, Tehran, Iran

Abstract

The integration of artificial intelligence and machine learning into chemical engineering represents a paradigm transformation that fundamentally reconceptualizes practice across molecular, process, and industrial scales. This comprehensive narrative review synthesizes recent advances in AI/ML methodologies, including graph neural networks, physics-informed neural networks, deep reinforcement learning, and generative artificial intelligence – and critically evaluates their applications spanning molecular property prediction, catalyst design, pharmaceutical development, reactor optimization, process control, and sustainability initiatives. Contemporary AI/ML approaches demonstrate unprecedented capabilities in navigating complex multi-scale phenomena while maintaining computational tractability and physical interpretability. Landmark achievements include 71% reduction in experimental iterations for reaction optimization through deep reinforcement learning, 98% accuracy in predictive maintenance using LSTM-based fault detection, sub-1% prediction errors in virtual metrology for semiconductor manufacturing, and substantial improvements in carbon capture efficiency through machine learning-guided materials discovery. Physics-informed neural networks address the critical challenge of plant-model mismatch by synergistically integrating mechanistic knowledge with data-driven learning, enabling extrapolation beyond training domains while respecting conservation laws. Explainable AI techniques, particularly SHAP analysis, enhance operational acceptance by 52% in safety-critical applications through transparent decision-making pathways. Despite remarkable progress, persistent challenges remain in data quality and standardization, model interpretability for regulatory compliance, computational scalability for real-time control, and integration with legacy industrial infrastructure. The review identifies transformative future directions including multi-modal learning frameworks, transfer learning for data-scarce applications, quantum machine learning for molecular design, and human-AI collaborative systems. Successful deployment demands interdisciplinary collaboration uniting chemical engineering domain expertise with computational intelligence, guided by principles of transparency, reproducibility, and responsible innovation to address sustainability imperatives while maintaining operational excellence and safety in chemical manufacturing.

* **Corresponding author, e-mail:** researchtahghigh6@gmail.com

Article info:

Received: 09 February 2026
Revised: 14 March 2026
Accepted: 27 May 2026

Keywords

Artificial Intelligence; machine learning; physics-informed neural networks; chemical process optimization; explainable AI

1. INTRODUCTION

Chemical engineering, a cornerstone of industrial innovation, has traditionally relied on first-principles modeling and empirical methods to design, optimize, and control chemical processes. However, the complexity and scale of modern chemical systems have pushed these classical approaches to their limits. In response, the integration of *Artificial Intelligence* (*AI*) and *Machine Learning* (*ML*) has emerged as a transformative development in the field. These data-driven tools are not only augmenting traditional modeling techniques but are increasingly enabling chemical engineers to perform tasks once thought too complex, non-linear, or data-intensive for conventional methods. *Artificial Intelligence* broadly refers to the simulation of human cognitive functions by machines, including reasoning, learning, and decision-making. Within *AI*, machine learning, especially supervised and unsupervised learning, has become particularly influential in scientific and engineering disciplines due to its ability to identify patterns and make predictions from large datasets. Chemical engineering, which generates vast quantities of experimental and operational data, is an ideal candidate for *AI* applications. *ML* models are being employed to accelerate reaction discovery, optimize reaction conditions, monitor process parameters, and enhance control strategies, all of which are central to chemical engineering practice. One of the earliest and most impactful areas where *AI* has shown promise is in reaction prediction. The ability to forecast reaction outcomes, yields, or even reaction pathways based on existing chemical data has traditionally required extensive domain expertise and trial-and-error experimentation. *AI*, particularly deep learning algorithms and neural networks, can be trained on large reaction databases to predict products, optimize reagents, and identify ideal conditions. This application is already accelerating drug development pipelines and green chemistry initiatives by enabling high-throughput screening and environmentally conscious process design (Mathur et al., 2021; Mondal 2024). These advancements are reducing reliance on energy-intensive laboratory experiments and minimizing waste, making the entire process more sustainable and cost-effective.

In addition to molecular design and synthesis, *AI* is also revolutionizing process modeling and optimization. Traditional process models, based on differential equations and thermodynamic principles, often struggle with highly non-linear or multi-scale systems. *ML* algorithms, including support vector machines and ensemble methods, have demonstrated superior predictive power in modeling complex chemical kinetics and multiphase flows, especially when high-quality historical or real-time data is available (Benavides-Hernández and Dumeignil, 2024; Dobbelaere et al. 2021). These models offer faster computation, adaptability to new data, and, when integrated with optimization algorithms, can guide engineers toward optimal operating conditions with greater accuracy and speed than manual tuning or heuristics. A significant evolution in *AI*-driven chemical engineering is its use in process control and automation. Advanced control systems are essential in ensuring product quality, minimizing energy use, and maintaining safety. Traditionally, model predictive control (*MPC*) systems have relied on deterministic process models. However, these can be computationally intensive and inflexible to dynamic process variations. *AI*-enabled controllers, on the other hand, can adapt in real-time, learn from sensor feedback, and even anticipate process upsets before they occur (Boskabadi et al., 2025). With developments in reinforcement learning and digital twin technologies, *AI*-based systems are not only optimizing operations but also acting as decision-support tools for plant operators and engineers. Despite its potential, the integration of *AI* into chemical engineering is not without challenges. Key concerns include the lack of interpretability of black-box models, the quality and quantity of training data, and the limited *AI* literacy among chemical engineers. Additionally, data silos, privacy issues, and computational costs remain barriers to widespread adoption. Yet, education initiatives and interdisciplinary collaboration are beginning to close this gap. Several universities have started integrating *AI* into their chemical engineering curricula, aiming to prepare a new generation of engineers equipped with both domain knowledge and computational skills (Bozal-Ginesta et al., 2025).

This review aims to explore the expanding role of *AI* in chemical engineering, from reaction prediction and synthesis to process modeling, optimization, and real-time control. By examining current applications, identifying challenges, and discussing future directions, we hope to offer a comprehensive overview of how *AI* is reshaping the practice and potential of chemical engineering.

It is important, however, to approach these advances with critical realism. Despite remarkable progress, the gap between benchmark performance and practical industrial deployment remains substantial and is frequently understated in the primary literature. Many *ML* models are validated on curated datasets under idealized conditions that poorly reflect the noise, non-stationarity, and incomplete observability of real industrial processes. The reproducibility crisis documented across computational sciences is acutely relevant to chemical engineering *AI*, where molecular diversity and process variability are intrinsic rather than correctable features. Moreover, the uncritical reporting of performance metrics such as R^2, *AUROC*, and *MAE* risks obscuring whether a model has genuinely learned physically meaningful representations or has instead overfit to statistical artifacts of a particular training dataset. Absent prospective validation, uncertainty quantification, and adversarial testing under distribution shift, high benchmark scores provide limited assurance of real-world reliability. This review therefore aims to provide not merely an inventory of what *AI* can do in chemical engineering, but a calibrated and critical assessment: acknowledging landmark achievements while explicitly identifying where evidence remains preliminary, where generalization claims are unsupported, and where foundational challenges have received insufficient attention from the community.

2. AI AND ML FUNDAMENTALS FOR CHEMICAL ENGINEERS

The rapid integration of *AI* and *ML* into chemical engineering requires a foundational understanding of how these systems operate, especially as they diverge from traditional mechanistic modeling paradigms. Unlike first-principles approaches, *ML* models infer patterns from empirical data without explicit physical rules. For chemical engineers, this represents a shift from deterministic frameworks to probabilistic, data-driven inference – offering flexibility, speed, and scalability but introducing challenges in interpretability and data dependency.

At the heart of modern *AI* applications in chemical engineering are supervised and unsupervised learning paradigms. Supervised learning algorithms, such as feedforward neural networks, support vector machines, and gradient-boosted trees, require labeled data to map input features to desired outputs. These models excel in tasks like reaction yield prediction, fault classification in reactors, and sensor calibration. For instance, Dobbelaere et al. (2021) demanded the applicability of supervised *ML* in process modeling, where traditional approaches struggled with non-linearities and stochastic perturbations in high-throughput operations.

Unsupervised learning, on the other hand, is pivotal in exploratory data analysis. Techniques like principal component analysis, k-means clustering, and autoencoders are utilized to uncover latent patterns in spectral data, reaction databases, or sensor arrays without predefined outcomes. These tools assist in anomaly detection, feedstock characterization, and optimization of complex multi-variate systems. Recent studies have shown that unsupervised models can cluster reaction types or operational regimes based on historical plant data, providing actionable insights without expert-coded thresholds (Wieder et al., 2020).

Surpassing traditional *QSAR* models in both performance and generalizability (Wieder et al. 2020). Their ability to capture long-range dependencies and contextual information

makes them well-suited to model multi-step synthesis, offering probabilistic assessments of alternative routes (Al Sharah et al., 2024). To maximize yield while avoiding process instabilities – tasks that are traditionally managed through rule-based *PID* or *MPC* systems (Hoque et al., 2024).

Despite their power, these models introduce fundamental concerns for chemical engineers, especially regarding model interpretability and trust. Most deep learning systems operate as "black boxes," offering high accuracy but little explanation. This opaqueness conflicts with the engineering imperative for transparent, auditable, and physically grounded decision-making. As such, there is growing interest in interpretable *ML*, including attention mechanisms, Shapley values, and symbolic regression, which attempt to bridge the gap between statistical learning and mechanistic understanding.

Furthermore, the successful deployment of *AI* systems is contingent on data quality – a challenge in chemical engineering, where experimental datasets are often sparse, noisy, and heterogeneous. Transfer learning and data augmentation techniques are increasingly used to overcome these limitations by leveraging pretrained models or generating synthetic data. Nonetheless, robust model validation remains essential, requiring cross-validation, out-of-sample testing, and domain-specific sanity checks to ensure reliability in real-world deployment.

A dimension of *AI*/*ML* in chemical engineering that merits more critical attention than it typically receives is the distinction between interpolation and extrapolation performance. Neural networks and gradient-boosted tree models are powerful interpolators within their training distribution but are notoriously unreliable extrapolators. In chemical engineering, where scale-up, novel feedstocks, and changing process conditions routinely push systems outside the envelope of historical data, this distinction is not academic – it determines whether a deployed model provides reliable guidance or dangerously misleading predictions.

Gaussian process regression offers principled uncertainty quantification that explicitly signals when predictions are made far from training data, though its computational scaling has traditionally limited applicability to large datasets. Conformal prediction methods offer distribution-free, coverage-guaranteed uncertainty bounds for any underlying *ML* model and are beginning to attract attention in pharmaceutical and materials discovery contexts.

For chemical process monitoring, where an *ML* model must reliably detect anomalous operating regimes, the ability to characterize prediction uncertainty under distribution shift is arguably more important than point-prediction accuracy. These considerations should be central to the validation methodology of any *ML* system intended for deployment in safety-critical chemical processes.

Indeed, for chemical engineers to harness the full potential of *AI*, a hybrid approach is needed – combining domain expertise with data-centric modeling. While *AI* tools extend beyond the limitations of traditional methods, their responsible application requires a deep appreciation of their statistical nature, assumptions, and failure modes. This necessitates a curriculum shift that integrates *AI* literacy into core chemical engineering education, as advocated by Wu et al. (2023), who showed that targeted *AI* instruction improves both student outcomes and confidence in applying these tools in complex systems. Critically, however, this integration must be pursued with care: *AI* tools should serve as supplements to – rather than substitutes for – rigorous training in chemical engineering fundamentals, ensuring graduates retain deep domain expertise alongside computational proficiency.

3. MOLECULAR AND MATERIALS SCALE APPLICATIONS

3.1. Molecular property prediction and structure-property relationships

The prediction of molecular properties through quantitative structure-property relationships (*QSPR*) represents a foundational application of *ML* in chemical engineering at the molecular scale (Chinta and Rengaswamy, 2019). Traditional *QSPR* approaches, dating to the 1960s, have evolved dramatically with modern *ML* algorithms that can capture complex nonlinear relationships between molecular structure and physicochemical properties.

Contemporary *QSPR* methodologies employ diverse molecular representations including molecular descriptors, fingerprints, and graph-based encodings to establish predictive models (Yan et al., 2020; Yang et al., 2019). Graph neural networks (*GNNs*), particularly directed message-passing neural networks (*D-MPNNs*), have emerged as state-of-the-art architectures for molecular property prediction, demonstrating superior performance across diverse chemical datasets (Heid et al., 2024).

The Chemprop package, implementing *D-MPNN* architecture, provides accessible infrastructure for predicting properties ranging from solubility and reactivity to absorption, distribution, metabolism, excretion, and toxicity (*ADMET*) characteristics. These models achieve remarkable accuracy by learning hierarchical molecular representations that capture both local atomic environments and global molecular topology (Heid et al., 2024; Wieder et al., 2020).

Recent advances address the critical challenge of data scarcity through transfer learning and multi-task learning frameworks. Adaptive checkpointing with specialization (*ACS*) mitigates negative transfer in imbalanced datasets while preserving beneficial inter-task knowledge sharing, enabling accurate property prediction with as few as 29 labeled samples (Eraqi et al., 2025).

This capability proves particularly valuable for specialized applications such as sustainable aviation fuel design, where limited experimental data constrains conventional modeling approaches. The integration of physics-based descriptors with data-driven learning, exemplified by fastprop and similar frameworks – combines computational efficiency with interpretability, achieving state-of-the-art performance while maintaining domain relevance (Burns and Green, 2025).

3.2. Catalyst design and discovery

Heterogeneous catalysis represents one of the most impactful domains for *ML* applications in chemical engineering, where the complexity gap between theoretical models and experimental reality has historically impeded rational catalyst design (Benavides-Hernández and Dumeignil, 2024; Goldsmith et al., 2018). Machine learning approaches address multiple aspects of this complexity, including configurational space exploration, reaction pathway enumeration, active site identification, and performance prediction (Table 1).

The integration of *ML* with high-throughput experimentation (*HTE*) and computational methods creates synergistic workflows that dramatically accelerate catalyst discovery cycles (Benavides-Hernández and Dumeignil, 2024; Goldsmith et al., 2018). Modern *ML* applications in heterogeneous catalysis encompass several critical categories. First, *ML* models predict adsorption energies and *d*-band centers that serve as descriptors for catalytic activity, substantially reducing the computational expense of density functional theory (*DFT*) calculations for complex surface chemistries (Mou et al., 2023).

Machine-learned force fields (*MLFFs*) enable molecular dynamics simulations at near-*DFT* accuracy with orders of magnitude reduction in computational cost, facilitating exploration of reaction mechanisms and free energy landscapes (Omranpour et al., 2025). These approaches prove particularly valuable for screening large catalyst libraries where exhaustive *DFT* calculations remain prohibitively expensive.

Graph neural networks trained on large-scale databases such as the Open Catalyst Project (*OC*20) demonstrate remarkable capability in predicting catalyst properties and reaction pathways, achieving energy prediction errors below 0.05 *eV* relative to *DFT* benchmarks (Bozal-Ginesta et al., 2025).

The emergence of interpretable *ML* methods – incorporating physics-inspired feature engineering, Bayesian statistical learning, and theory-infused architectures – addresses the critical challenge of model explainability while maintaining predictive accuracy (Xin et al., 2023). These interpretable frameworks enable extraction of design principles and mechanistic insights that guide experimental catalyst development, bridging the gap between data-driven predictions and domain understanding.

3.3. Drug discovery and pharmaceutical engineering

ADMET property prediction has become a cornerstone application of *ML* in pharmaceutical development, addressing the critical bottleneck that contributes to high attrition rates in drug discovery pipelines (Terstappen and Reggiani, 2001; Vora et al., 2023). Modern *ML* platforms such as *ADMET-AI*,

Table 1. Representative ML applications in heterogeneous catalyst design.

Application domain	ML approach	Key achievement	Accuracy/Performance	Reference
Adsorption energy prediction	Graph Neural Networks	Universal binding energy models for multimetallic surfaces	$MAE < 0.15$ eV	Mou et al., 2023
Surface reaction mechanisms	Machine-Learned Force Fields	Automated reaction pathway exploration	$\Delta E_{error} \sim 0.05$ eV vs. *DFT*	Omranpour et al., 2025
Active site identification	Bayesian Optimization	Accelerated screening of single-atom catalysts	Speedup: 10–100× vs. *DFT*; $MAE < 0.1$ eV	Xin et al., 2023
Catalyst stability prediction	Gradient Boosting Regression	Thermal and chemical stability assessment	$R^2 > 0.85$	Goldsmith et al., 2018
Oxygen reduction reaction	Explainable AI (SHAP)	Descriptor importance ranking for ORR activity	$R^2 > 0.80$; experimental validation	Xin et al., 2023
CO_2 reduction	Transfer Learning	Performance prediction with limited data	$R^2 = 0.80$–0.90; 15–30% improvement	Bozal-Ginesta et al., 2025

HelixADMET, and commercial systems provide rapid, accurate predictions across 40 + *ADMET* endpoints, enabling efficient filtering of large chemical libraries generated through virtual screening and generative *AI* approaches (Swanson et al., 2024; Zhang et al., 2022).

These platforms employ sophisticated graph neural network architectures that achieve competitive performance on standardized benchmarks while maintaining computational efficiency sufficient for screening thousands of compounds in minutes.

The integration of *ML* into drug discovery workflows spans multiple stages, from hit identification through lead optimization. Multi-task deep neural networks simultaneously predict multiple *ADMET* properties, exploiting correlations among endpoints to improve predictive accuracy while reducing computational redundancy (Wenzel et al., 2019).

Transfer learning strategies prove particularly valuable for rare toxicity endpoints where limited training data is available, though care must be exercised to avoid negative transfer when applying models to narrow therapeutic areas or specialized compound classes (Niazi, 2025). Recent developments in quantum-informed molecular representations and transformer-based architectures further enhance prediction accuracy by capturing electronic structure information and long-range molecular interactions (Kim et al., 2024).

The pharmaceutical industry's adoption of in silico *ADMET* modeling reflects its demonstrated value in reducing late-stage failures and accelerating development timelines. The global in silico drug discovery market, valued at $ 3.61 billion in 2024, projects growth to $ 7.22 billion by 2030, driven by the imperative to "fail early, fail cheap" given that bringing a new molecular entity to market surpasses $ 2.6 billion (Pathan et al., 2025).

Regulatory acceptance of computational methods, exemplified by the *FDA*'s 2025 decision to phase out mandatory animal testing for many drug types, signals a paradigm shift toward increased reliance on validated in silico approaches for safety and efficacy assessment (Table 2).

Typical *ADMET* prediction accuracy ranges:

- Solubility: $R^2 = 0.75$–0.85
- Blood-brain barrier: $AUROC = 0.85$–0.92
- hERG blocking: $AUROC = 0.80$–0.88
- CYP inhibition: $AUROC = 0.82$–0.90
- Oral bioavailability: $AUROC = 0.75$–0.82

3.4. Novel materials development

3.4.1. Polymer design and characterization

Machine learning has catalyzed a paradigm shift in polymer materials research, enabling the systematic exploration of vast chemical design spaces that were previously inaccessible through traditional trial-and-error approaches (Gao et al., 2024; Yue et al., 2025). The *ML*-assisted polymer design workflow encompasses database construction, molecular representation, property prediction model development, virtual screening, and experimental validation.

This structured framework addresses the unique challenges posed by polymers' structural complexity, multiscale architecture, and limited availability of high-quality training data (Cao et al., 2026). Transfer learning and multitask learning strategies have emerged as essential techniques for overcoming data scarcity in specialized polymer applications.

Notable successes include the discovery of polymers with high thermal conductivity, where transfer learning from large databases (PoLyInfo, QM9) to limited thermal property data enabled identification and experimental validation of novel high-performance materials (Wu et al., 2019). Bayesian molecular design frameworks integrate uncertainty quantification with sequential optimization, guiding efficient exploration of chemical space while balancing exploitation of known favorable regions with exploration of uncharted territories.

Recent applications demonstrate *ML*'s transformative potential across diverse polymer domains. In membrane separation, *ML*-driven inverse design accelerates identification

Table 2. Performance metrics of state-of-the-art ADMET prediction platforms.

Platform	Architecture	Number of endpoints	Prediction speed	Key features	Reference
ADMET-AI	Chemprop-RDKit GNN	41	Fastest (45% faster)	DrugBank comparison	Swanson et al., 2024
HelixADMET	Self-supervised NNs	Multiple	High-throughput	Knowledge transfer	Zhang et al., 2022
ADMET Predictor	Multi-algorithm ensemble	175 +	Commercial-grade	Integrated PBPK	Venkataraman et al., 2025
PharmaBench	Large Language Models	11 datasets, 52 K entries	Benchmark-focused	LLM curation	Niu et al., 2024
Deep-PK/DeepTox	Graph-based multitask	PK and toxicity	Platform-dependent	Graph descriptors	Pathan et al., 2025

of polymeric materials with optimized permeability-selectivity trade-offs for gas separation and water treatment applications (Dangayach et al., 2025).

For energy storage, *AI*-guided discovery of dielectric polymers achieved simultaneous high energy density and thermal stability – properties traditionally considered mutually exclusive – through systematic exploration of polynorbornene and polyimide chemical spaces (Georgia Institute of Technology, 2024). Flame-retardant polymer discovery leverages operando analytical techniques integrated with virtual combustion generators to capture intermediate combustion information, enabling more accurate property prediction and rational design of fire-safe materials (Wang et al., 2024a).

3.4.2. Deep eutectic solvents and ionic liquids

Designer solvents, particularly deep eutectic solvents (*DES*) and ionic liquids (*ILs*), represent an emerging class of tunable materials where *ML* approaches accelerate property prediction and formulation optimization. Group contribution (*GC*) methods combined with *ML* algorithms predict temperature-dependent properties including viscosity, surface tension, thermal conductivity, and cytotoxicity with remarkable accuracy when appropriate data splitting schemes and algorithms are employed (Cao et al., 2024).

The critical challenge lies in capturing the complex non-additive interactions between cation-anion pairs or hydrogen bond donor-acceptor components that determine solvent properties.

Machine learning models for *IL* and *DES* properties must account for temperature dependence, ionic strength effects, and compositional variations across multicomponent systems. Point-based versus *IL*-based dataset splitting significantly impacts model performance for temperature-dependent properties, highlighting the importance of rigorous cross-validation strategies (Cao et al., 2024).

Random forest and extreme gradient boosting algorithms frequently emerge as optimal choices for *IL* property prediction, balancing predictive accuracy with computational efficiency. The integration of quantum mechanical descriptors and molecular dynamics-derived features further enhances model performance, particularly for properties governed by electronic structure and intermolecular interactions.

3.4.3. Nanomaterials and advanced functional materials

The extension of *ML* methodologies to nanomaterial design addresses unique challenges related to size-dependent properties, surface effects, and synthesis-structure relationships. Generative models create virtual libraries of nanoparticle compositions and morphologies, which are subsequently screened using *ML* property predictors trained on high-throughput computational or experimental data (Bozal-Ginesta et al., 2025). This approach has accelerated discovery of catalytic nanoparticles, quantum dots with tailored optical properties, and functional nanocomposites for energy conversion applications.

Integration of *ML* with autonomous experimental platforms – self-driving laboratories – enables closed-loop optimization where experimental results continuously refine predictive models and guide subsequent synthesis attempts. These cyber-physical systems combine robotic synthesis, in-line characterization, and adaptive algorithms to achieve accelerated materials discovery cycles measured in days rather than years (Boskabadi et al., 2025). The success of such platforms depends critically on developing robust models that generalize beyond training data while accounting for experimental variability and synthesis-property relationships.

4. REACTION AND PROCESS UNIT SCALE

4.1. Chemical reaction optimization

Deep reinforcement learning (*DRL*) has emerged as a transformative methodology for autonomous chemical reaction optimization, addressing the fundamental challenge of navigating vast experimental spaces efficiently. The Deep Reaction Optimizer (*DRO*) framework exemplifies this paradigm, employing modified long short-term memory architectures to iteratively record reaction outcomes and select new experimental conditions that maximize target objectives such as yield or selectivity (Zhou et al., 2017).

This approach demonstrated superior performance over conventional blackbox optimization algorithms, requiring 71% fewer experimental iterations while achieving optimal conditions in as little as 30 minutes when integrated with accelerated microdroplet reaction platforms. The stochastic exploration strategy incorporated within *DRL* frameworks mitigates the greedy policy limitation that causes entrapment in local optima, systematically balancing exploitation of promising conditions with exploration of uncharted parameter spaces (Zhou et al., 2017).

The *RE-EXPLORE* framework advances reaction optimization by integrating *DRL* with recurrent neural network-based generative models to identify novel reactants and catalysts that maximize yield or selectivity (Hoque et al., 2024). Pretrained on large chemical databases containing hundreds of thousands to blending millions of molecules, these generative models are enriched with diverse chemical knowledge that informs the search process.

The engineered reward function incorporates Tanimoto-based uniqueness factors within the reinforcement learning loop, addressing the critical challenge whereby standard *RL* methods prioritize immediate exploitation, resulting in repetitive generation of chemically similar molecules. This architectural

innovation enables efficient exploration of chemical space while maintaining focus on performance optimization objectives. The integration of user-defined core fragments further facilitates learning of specific reaction types, providing chemists with control over the molecular scaffolds explored during optimization (Hoque et al., 2024).

Multi-objective reinforcement learning extends single-objective optimization to address the realistic complexity of pharmaceutical lead optimization, where multiple competing objectives such as potency, selectivity, pharmacokinetics, and synthetic accessibility must be simultaneously optimized (Zhou et al., 2017). The *MolDQN* framework demonstrates this capability through direct molecular modifications that ensure 100% chemical validity while maximizing drug-likeness and maintaining structural similarity to lead compounds.

Recent extensions to catalytic reaction mechanism elucidation leverage high-throughput deep reinforcement learning integrated with first-principles calculations to autonomously predict reaction pathways with near-*DFT* accuracy at dramatically reduced computational cost (Lan et al., 2024). This reaction-agnostic framework demonstrates excellent generalizability across diverse catalytic systems, enabling efficient mechanistic investigations that traditionally required extensive human expertise and computational resources.

Despite impressive quantitative results, several important limitations of *DRL*-based reaction optimization deserve explicit discussion. First, the reward function design problem is nontrivial and directly determines what the *RL* agent optimizes. In pharmaceutical synthesis, the chemist's true objective involves multiple competing criteria including yield, selectivity, atom economy, reagent safety, solvent greenness, and scalability to manufacturing conditions.

Collapsing this multidimensional objective into a scalar reward inevitably privileges some criteria at the expense of others, and small changes in reward formulation can dramatically alter recommended conditions. Second, the structural coverage of pretrained generative models underlying frameworks such as *RE-EXPLORE* is heavily biased toward motifs dominating publicly available reaction databases such as *USPTO* and Reaxys, which are skewed toward medicinal chemistry.

This means *RL*-based methods may fail to generate genuinely novel molecular frameworks or conditions outside the represented chemistry – a form of extrapolation failure difficult to detect from within-distribution metrics. Third, integration of *DRL* with physical experimentation raises practical concerns: unlike simulation environments where failed experiments carry no cost, real flow chemistry platforms impose throughput limits, reagent constraints, and safety boundaries that must be explicitly encoded in the action space and constraint handling of the *RL* agent. These implementation details, critical to real-world deployment, are frequently minimized in published reports focused primarily on algorithmic performance.

4.2. Reactor design and modeling

Computational fluid dynamics (*CFD*) has evolved from a research tool to an indispensable component of modern reactor design and optimization, providing detailed insights into flow patterns, mixing behavior, residence time distributions, temperature gradients, and reaction kinetics within complex reactor geometries (Lodh, 2024). The integration of *CFD* with machine learning accelerates model development and enables real-time optimization capabilities previously unattainable with traditional computational approaches.

CFD simulations optimize stirred-tank reactor configurations by evaluating different impeller designs, agitation speeds, and baffle arrangements to achieve uniform mixing, maximize reaction rates, and minimize stagnant zones that degrade product quality and yield. A critical and frequently underappreciated limitation in *CFD–ML* integration is the turbulence closure problem. Conventional *CFD* relies on Reynolds-Averaged Navier–Stokes (*RANS*) models – most commonly $k-\varepsilon$ and $k-\omega SST$ variants – whose empirical constants were calibrated against canonical flows such as flat plates and pipe flows.

These models are well-documented to perform poorly in geometrically complex reactor environments with strong swirl, recirculation zones, and reactive multiphase interactions. Machine learning offers a physically motivated path to improved closures: data-driven models trained on high-fidelity Direct Numerical Simulation (*DNS*) or Large Eddy Simulation (*LES*) datasets can learn nonlinear stress–strain relationships that classical linear eddy-viscosity models are fundamentally incapable of representing.

However, *ML* turbulence closures trained on specific flow configurations may propagate prediction errors through coupled governing equations in ways that are complex to detect, particularly when deployed on reactor geometries outside their training distribution. Critically, a solution may satisfy numerical convergence criteria while producing physically incorrect species concentration fields and temperature profiles – a failure mode invisible to standard residual-based convergence monitoring.

This underscores the necessity of embedding hard physical constraints – realizability conditions for the Reynolds stress tensor, frame-invariance, and conservation law compliance – directly into any *ML* model augmenting *CFD* closure terms, rather than relying on soft loss penalization. For reactive flow systems – including tubular reformers, fluidized bed reactors, and gas–liquid bubble columns – the coupling between fluid mechanics, heat transfer, mass transfer, and chemical kinetics creates multiphysics complexity that exposes a foundational tension in *CFD–ML* surrogate strategies.

In steam methane reforming reactors, steep radial temperature gradients near tube walls, catalyst activity gradients along the bed, and pressure-dependent thermodynamic equilibrium of simultaneous endothermic reactions create a system

where errors in the predicted flow field propagate nonlinearly into conversion and selectivity predictions. *ML* surrogate models trained on isolated physical subsystems may achieve high individual accuracy while failing substantially when interconnected, because coupling terms introduce input variables absent from any single training distribution.

The most physically defensible approach involves operator learning frameworks – such as *DeepONet* or Fourier Neural Operators – that learn solution mappings for fully coupled *PDE* systems. However, their application to industrially complex reactive *CFD* remains at an early stage and awaits rigorous prospective validation. Scale-up further compounds these challenges: turbulence intensity, interfacial area-to-volume ratios, and axial dispersion all change non-trivially from bench to pilot to industrial scale, meaning *ML* surrogates validated at one scale cannot be reliably transferred to another without retraining on scale-specific simulation data – significantly diminishing the claimed cost savings of the surrogate approach.

Digital twins represent the apex of *CFD* integration into process engineering, creating virtual replicas of physical reactors that synchronize with real-time operational data to enable predictive control and optimization (De Carfort et al., 2024; Lodh, 2024). These cyber-physical systems combine mechanistic *CFD* models with data-driven machine learning algorithms to capture both fundamental physics and empirical relationships learned from operational history.

Samsung Biologics' bioprocess digital twin exemplifies this hybrid modeling approach, integrating advanced *CFD* with mechanistic cell kinetics models and deep learning systems to deliver highly accurate predictions tailored to individual biopharmaceutical products (Lodh, 2024). Unlike purely data-driven digital twins, mechanistic integration ensures physical consistency and enables extrapolation beyond training data, critical capabilities for managing process variations and optimizing novel production campaigns.

The computational demands of *CFD* traditionally constrained real-time applications, limiting digital twin implementations to offline analysis and periodic model updates. Recent methodological advances address this bottleneck through automatic generation of *CFD*-based compartment models that convert detailed three-dimensional flow field solutions into simplified systems of mass balance equations (De Carfort et al., 2024).

This reduced-order modeling approach preserves essential mixing characteristics and transport phenomena while achieving computational speeds faster than real-time, essential for predictive control applications. Validation studies demonstrate that *CFD*-derived compartment models accurately replicate tracer mixing profiles and, when coupled with biokinetic models, correctly predict the impact of substrate gradients on fermentation performance across scales from bench to production (De Carfort et al., 2024).

Artificial intelligence-enhanced *CFD* workflows further accelerate reactor optimization through surrogate modeling and automated design space exploration. Machine learning models trained on *CFD* simulation databases predict key performance indicators such as mixing time, heat transfer coefficients, and conversion efficiency at orders of magnitude faster than full *CFD* solutions, enabling rapid screening of thousands of design variants (Lodh, 2024).

The integration of physics-informed neural networks with *CFD* creates hybrid models that respect conservation laws while learning complex closure relationships from data, addressing the persistent challenge of turbulence modeling in reactive multiphase flows (Benavides-Hernández and Dumeignil, 2024). These advanced capabilities position *AI*-enhanced *CFD* as the foundation for next-generation autonomous reactor design systems that optimize performance while accounting for manufacturing constraints, safety requirements, and economic objectives simultaneously.

5. PLANT-LEVEL AND INDUSTRIAL APPLICATIONS

5.1. Process control and monitoring

Physics-informed neural networks have revolutionized model predictive control by enabling real-time optimization of nonlinear chemical processes while maintaining computational tractability (Patel et al., 2024). The *PINN*-based *MPC* framework integrates governing physical equations into the neural network loss function, creating hybrid models that leverage both approximate mechanistic knowledge and noisy measurement data.

When applied to solid oxide fuel cells, autocatalytic systems, and plug flow reactors, *PINN*-based *MPC* demonstrated superior performance compared to purely data-driven models, particularly in extrapolation scenarios where training data coverage was limited (Patel et al., 2024). The framework achieved real-time feasibility while maintaining prediction accuracy, addressing the long-standing trade-off between model complexity and computational cost that has constrained traditional *MPC* implementations.

Input convex neural networks further advance explicit *MPC* by formulating control optimization as convex problems, dramatically reducing computational burden while ensuring global optimality (Wang et al., 2024b). Unlike conventional neural network-based *MPC* that requires solving nonconvex optimization problems at each time step, *ICNN-MPC* converts the control problem into mixed-integer quadratic programming that can be solved orders of magnitude faster.

This computational efficiency proves critical for fast dynamic processes where millisecond-scale decision making is required. Industrial implementations demonstrate that *LSTM* and artificial neural network-based *MPC* models outperform traditional reduced-order approaches, achieving superior R^2 and *RMSE* metrics while maintaining computational speeds compatible with real-time control (Dan et al., 2025).

A critical assessment of *AI*-based process control deployments reveals that published performance metrics are frequently obtained under conditions substantially more favorable than those in routine industrial operation. Most reported *LSTM-MPC* and *PINN-MPC* benchmarks use simulation environments or laboratory-scale equipment with high sensor density, low noise, and stable disturbances. Real industrial processes involve intermittent sensor faults, actuator saturation, valve hysteresis, and feed composition disturbances that are complex to represent faithfully in training data.

The closed-loop stability guarantees available for linear *MPC* with terminal cost formulations do not extend straightforwardly to neural network-based controllers, whose relationship between network parameters and closed-loop behavior is opaque. Lyapunov-based stability certificates for neural network controllers require restrictive structural assumptions limiting practical applicability.

Perhaps the most underappreciated challenge is process nonstationarity: equipment aging, catalyst deactivation, and seasonal feedstock variation continuously shift the dynamics the *ML* controller was trained to handle, requiring either continuous online retraining – raising safety certification concerns – or periodic batch retraining with careful version management.

Regulatory frameworks for *AI*-based control in safety-critical chemical manufacturing, including COMAH sites in Europe and PSM-covered facilities in the United States, do not yet provide clear guidance on validation and change management requirements for adaptive *ML* controllers, creating a barrier to industrial adoption that technical performance alone cannot overcome.

5.2. Soft sensors and virtual metrology

Virtual metrology represents a critical enabler of Industry 4.0 manufacturing, providing cost-effective alternatives to physical measurements while enabling wafer-to-wafer control and early drift detection (Gentner et al., 2021; He and Wang, 2011). Deep learning architectures, particularly one-dimensional convolutional neural networks combined with bidirectional *LSTM* networks and attention mechanisms, have demonstrated exceptional capability in virtual metrology applications for semiconductor chemical mechanical planarization (*CMP*) (Breidung et al., 2025).

These architectures automatically extract relevant features from raw time-series sensor data, eliminating the need for manual feature engineering while achieving prediction errors below 1% for material removal rate estimation. The incorporation of attention mechanisms enables the model to focus on critical temporal patterns that most strongly influence product quality.

Hybrid virtual metrology frameworks that integrate physics-based models with machine learning achieve robust predictions across equipment variations and process conditions (Deivendran et al., 2025; Rahman et al., 2024). For *CMP* applications, physics-based models estimate critical but unmeasurable parameters such as slurry temperature and mean abrasive particle size from available sensor data, which then serve as inputs to *ML* soft sensors for material removal rate prediction.

This hybrid approach ensures physical consistency while leveraging *ML*'s capability to capture complex empirical relationships not captured by simplified mechanistic models. Domain adaptation techniques enable virtual metrology models trained on one equipment set to transfer effectively to identical-in-design systems with different data distributions, addressing the critical challenge of model scalability in multi-equipment manufacturing environments (Breidung et al., 2025) (Table 3).

The interpretability of virtual metrology models has emerged as a critical requirement for industrial acceptance and regulatory compliance. *SHAP* analysis applied to XGBoost and random forest models reveals that polishing time and carrier rotation speed are the most influential features for *CMP* non-uniformity prediction, with spatial material removal patterns exhibiting strong dependencies on local process conditions (Deivendran et al., 2025).

Table 3. Virtual metrology and soft sensor applications in industrial processes.

Application Domain	ML Architecture	Key Performance Metric	Industrial Impact	Reference
CMP Material Removal Rate	1*DCNN-BiLSTM* with Attention	$MAE < 1\%$	Wafer-to-wafer control	Breidung et al., 2025
Semiconductor *VM* (Multi-process)	Deep Learning Domain Adaptation	Model transfer	Scalable *VM* deployment	Breidung et al., 2025
Bioprocess Control	*LSTM*-based *MPC*	$R^2 > 0.90$	Maximizes product titer	Dan et al., 2025
Sugar Crystallization	*LSTM* Soft Sensor	2.12% residence reduction	Energy and CO_2 savings	Boskabadi et al., 2025
Chemical Vapor Deposition	*RBF* Neural Network	$MAPE = 0.34\%$, Max error $= 1.15\%$	Early drift detection	He and Wang, 2011
Plasma Etch Process	*RNN* with *POD*	Accurate profile prediction	Cycle time reduction	Xiao et al., 2021

This interpretability facilitates root cause analysis when quality deviations occur and guides process engineers in making informed adjustments to operational parameters. Advanced virtual sensor frameworks for batch crystallization processes demonstrate that *LSTM*-based soft sensors can reduce mean residence time by 2.12%, consequently decreasing energy consumption, CO_2 footprint, and increasing production capacity (Boskabadi et al., 2025).

5.3. Fault detection and predictive maintenance

AI-powered predictive maintenance has emerged as a high-value, low-risk application domain where chemical engineering companies achieve substantial return on investment (Xiao et al., 2021). T^2-*LSTM* frameworks for motor fault detection in chemical plants achieve 98% agreement between predicted and actual anomalies over multi-month operational periods, demonstrating practical viability for critical rotating equipment (Wang, 2024).

The Hotelling T^2 index calculated from triaxial vibration measurements provides a comprehensive health indicator, while *LSTM* trend modeling establishes adaptive warning thresholds that account for gradual equipment degradation patterns. This approach enables early intervention before catastrophic failures occur, preventing costly unplanned shutdowns and safety incidents.

Gaussian process regression with automatic kernel selection provides an advanced framework for fouling prediction in heat exchangers and distillation columns, enabling transition from reactive to prescriptive maintenance strategies (Negri et al., 2025). The *P-F* curve framework illustrates how predictive maintenance progressively anticipates failure events, with *AI*-based approaches detecting anomalies weeks before mechanical failure would manifest through traditional monitoring.

Companies like BASF and Dow report that *AI*-driven predictive maintenance for pumps, compressors, and heat exchangers detects minute vibration pattern anomalies weeks in advance, resulting in substantial cost savings and avoidance of production shutdowns (Negri et al., 2025).

Explainable *AI* techniques, particularly *SHAP* applied to adversarial autoencoders, enable transparent fault diagnosis in complex chemical processes (Sanjuan Gómez, 2021). The framework generates health indicators that capture overall plant condition, with *SHAP* attributing any significant fluctuation to specific process variables and fault types.

When tested on continuous stirred-tank reactors and the Tennessee Eastman Process, the explainable *AI* framework diagnosed catalyst decay, heat transfer fouling, feed disturbances, sensor faults, and valve stiction more accurately than previous methods. The transparency provided by explainable *AI* proves essential for building operator trust and meeting regulatory requirements in safety-critical applications.

6. SUSTAINABILITY AND ENVIRONMENTAL APPLICATIONS

Machine learning has emerged as a critical enabler for addressing sustainability imperatives in chemical engineering, particularly in carbon capture, waste valorization, and circular economy implementation ((Manikandan et al., 2025)). The integration of high-throughput computational screening with data-driven modeling, termed the *SMART* (Scalable Modeling, Artificial Intelligence, and Rapid Theoretical calculations) approach, combines expertise from chemical, materials, environmental, and data science to advance carbon capture technologies (Lei et al., 2023).

Multi-scale computational design frameworks employing genetic algorithms and machine learning identify metal-organic frameworks optimized for postcombustion CO_2 capture, with process-level modeling predicting performance in modified Skarstrom cycles (Deng and Sarkisov, 2024). These frameworks discover materials exhibiting high CO_2/N_2 selectivity through either size-exclusion mechanisms in small pores or multiple interaction sites in larger pores.

AI-driven optimization of carbon capture processes delivers substantial techno-economic and environmental benefits critical for scaling technologies to meet net-zero emissions targets. Machine learning models reduce capital and operational expenditures by optimizing process parameters and accelerating sorbent material discovery, addressing cost barriers that have historically impeded widespread adoption (Manikandan et al., 2025; Smart et al., 2025).

The Process-Informed design of tailor-made Sorbent Materials tool exemplifies this approach, using simulations and machine learning to identify cost-effective, sustainable material-process combinations while simultaneously evaluating economic and environmental impacts during the design phase rather than as post-hoc considerations. Artificial intelligence integration into wastewater treatment carbon capture systems demonstrates 97% accuracy in predicting CO_2 emissions and 95% accuracy for H_2S emissions, enabling real-time process optimization that reduces greenhouse gas emissions while recovering valuable resources.

Beyond carbon capture, machine learning is being applied across a wider spectrum of environmental challenges in chemical engineering, though the depth of critical validation varies greatly across application areas. In wastewater treatment, *ML*-based models have been applied to predict effluent quality, optimize coagulant dosing, and detect process upsets in activated sludge systems. Gradient boosting and *LSTM* architectures have demonstrated prediction accuracy above 90% for biochemical oxygen demand (*BOD*) and chemical oxygen demand (*COD*) under controlled experimental conditions.

However, a recurring weakness in this literature is the short evaluation boxes used for model validation: performance reported over weeks or months fails to capture seasonal feed composition shifts, microbial community evolution, and equipment aging effects that determine long-term operational reli-

ability. For life cycle assessment (*LCA*) and techno-economic analysis (*TEA*), *ML* models are increasingly used to fill data gaps and accelerate scenario screening.

Graph neural network-based *LCA* frameworks can propagate uncertainty across thousands of process configurations in minutes. Nevertheless, the quality of *ML*-accelerated *LCA* depends entirely on the completeness and geographic representativeness of the underlying inventory databases, which remain highly uneven across industrial sectors. The risk of perpetuating or amplifying existing data biases in sustainability assessments through *ML* automation deserves explicit attention from the research community.

The circular economy represents an emerging and particularly challenging application domain for *AI* in chemical engineering. Predicting the recyclability of complex polymer blends, optimizing chemical depolymerization pathways for mixed plastic waste, and designing processes for critical metal recovery from electronic waste all require *ML* models capable of handling highly heterogeneous input streams with variable composition, contamination, and degradation states – conditions fundamentally different from the homogeneous, well-characterized feedstocks on which most published models are trained.

Generative *AI* and reinforcement learning show conceptual promise for designing recycling-compatible polymer architectures and optimizing separation sequences for mixed waste streams, but prospective experimental validation of computationally generated recommendations remains rare. This gap between in silico promise and demonstrated experimental utility is the central unresolved challenge for *AI* in circular economy applications. Honest reporting of this gap is essential if the field is to channel resources productively toward tractable near-term opportunities rather than aspirational applications requiring foundational advances that have not yet been achieved.

7. PHYSICS-INFORMED AND HYBRID APPROACHES

Physics-informed neural networks (*PINNs*) address the fundamental challenge of plant-model mismatch in process systems by leveraging first-principles knowledge alongside data-driven techniques (Benavides-Hernández and Dumeignil, 2024; Moayedi et al., 2024). *PINNs* respect physical characteristics while yielding accurate dynamic models from process data, outperforming purely data-driven approaches like recurrent neural networks in predictive capability, particularly for extrapolation beyond training domains.

Applications to continuously stirred tank reactors and liquid-liquid separators demonstrate that *PINNs* infer immeasurable states with reasonable accuracy even when constitutive equations are unknown or incomplete (Moayedi et al., 2024; Velioglu et al., 2025). The decoupling-coupling training framework represents a methodological advance for modeling chemical reactor systems with multiphysics coupling (Wu et al., 2024). This approach decomposes complex physical domains into subdomains of fluid flow, heat transfer, and mass transfer combined with reaction kinetics, with each subdomain represented by specialized neural networks. Decoupling pre-training provides coarse but reasonable parameter distributions for initializing sub-networks before multiphysics coupling training, addressing the optimization challenges that arise when *PINNs* struggle with multiple coupled physical phenomena.

Physics-informed recurrent neural networks deployed with hyperparameter optimization enable modeling of dynamic processes by embedding physics through multi-objective loss functions or hybrid cell architectures performing Euler discretization (Asrav et al., 2023). These approaches achieve improved test performance and more robust architectures compared to purely data-driven alternatives, demonstrating particular value in the small-data regime common to chemical engineering applications.

Despite their theoretical appeal, physics-informed neural networks face a set of practical challenges that are frequently underemphasized in the literature. The primary difficulty lies in the formulation and weighting of the composite loss function, which must simultaneously minimize data mismatch, satisfy governing differential equations across the domain interior, and enforce boundary and initial conditions.

The relative weights assigned to these loss terms profoundly affect convergence behavior and solution quality, yet no principled, physics-agnostic method for selecting these weights currently exists. The *PINN* loss landscape is known to exhibit pathological behavior including saddle points, flat regions, and competing gradient magnitudes that cause training to stall or converge to physically spurious solutions.

For stiff chemical kinetics – where species concentrations evolve on timescales spanning many orders of magnitude, as in combustion or enzymatic reaction networks – standard gradient descent optimizers routinely fail to converge to physically correct solutions without problem-specific architectural modifications. Furthermore, published *PINN* validation studies in chemical engineering predominantly use forward problems with known analytical or numerical solutions as ground truth.

Inverse problems – inferring model parameters or unmeasured states from limited sensor data under realistic noise – represent the practically critical application, yet robust benchmarking of *PINN* performance on ill-posed inverse problems with proper uncertainty quantification remains sparse. Hybrid modeling – combining mechanistic first-principles models with data-driven components – is widely advocated as the pragmatic middle path between interpretable but inflexible white-box models and accurate but opaque black-box models.

The serial hybrid architecture, in which a mechanistic model generates primary state estimates subsequently corrected by an *ML* residual model, is particularly suited to chemical engineering where mass and energy balances are known but kinetic parameters are uncertain. This approach has been applied to fed-batch bioreactor modeling, distillation col-

umn optimization, and heat exchanger fouling prediction with demonstrated improvements in extrapolation accuracy.

The parallel hybrid architecture, in which mechanistic and *ML* components operate simultaneously and outputs are combined by a learned weighting function, offers greater flexibility but risks mode collapse where one component dominates. A critical and underreported limitation of both architectures is the error attribution problem: when a hybrid model fails in deployment, diagnosing whether the failure originated in the mechanistic component (wrong physical assumptions), the data-driven component (distribution shift), or the coupling mechanism (incorrect weighting) is non-trivial and often requires experiments specifically designed to isolate each component – a requirement rarely met in industrial deployment scenarios (Asrav et al., 2023; Wu et al., 2024).

8. EMERGING FRONTIERS

8.1. Explainable AI and interpretability

Explainable artificial intelligence (*AI*) has transitioned from academic curiosity to operational necessity in chemical engineering, with *SHAP* analysis increasing operational acceptance by 52% in safety-critical settings by visualizing neural network decision pathways (Romagnoli et al., 2024). Model-agnostic post-hoc interpretability methods enable identification of dominant material descriptors and feature contributions, building trust in *AI* predictions for high-stakes applications (Oviedo et al., 2022).

The *XpertAI* framework integrates *SHAP* with large language models to generate accessible natural language explanations of structure-property relationships, addressing the limitation that traditional *XAI* methods target technically oriented users rather than domain experts (Wellawatte and Schwaller, 2025). Applications to molecular property prediction, catalysis optimization, and materials design demonstrate that *SHAP* values reveal physically meaningful contributions.

For instance, *SHAP* analysis of *Pt* monolayer core-shell catalysts for oxygen reduction reactions voids the critical role of adsorbate resonance energies in chemical bonding, providing interpretable pathways for rational catalyst design (Omidvar et al., 2024). Regulatory agencies increasingly mandate transparency and accountability in *AI* systems supporting pharmaceutical development and chemical manufacturing.

The *FDA*'s 2025 draft guidance emphasizes traceable, context-aware validation throughout *AI* pipelines, positioning explainability not as luxury but prerequisite for deployment (Lavecchia, 2025). Hybrid approaches combining physics-based models with machine learning reduce prediction errors by 38–52% in catalysis simulations compared to pure data-driven methods while enhancing interpretability through incorporation of domain knowledge (Romagnoli et al., 2024).

This synergy proves particularly valuable in safety-critical reactor control and process optimization where understanding decision rationale enables rapid troubleshooting and regulatory compliance.

8.2. Digital twins and autonomous systems

Digital twin technology represents the convergence of mechanistic modeling, computational fluid dynamics (*CFD*), and machine learning to create virtual replicas that synchronize with real-time operational data (Lodh, 2024). Unlike purely data-driven surrogates, mechanistic integration ensures physical consistency and enables extrapolation beyond training data, essential capabilities for managing process variations and optimizing novel production campaigns.

Bioprocess digital twins demonstrate this integration by combining *CFD* with cell kinetics models and deep learning to deliver product-specific predictions, facilitating predictive control that maximizes product titer while respecting equipment constraints.

Despite transformative potential, *AI*/*ML* integration into chemical engineering confronts critical challenges that must be systematically addressed. Data quality, availability, and standardization remain persistent obstacles, with industrial plants collecting 12–15 sensor data types yet achieving interoperability in fewer than 40% of cases (Romagnoli et al., 2024). This fragmentation necessitates adaptive data pipelines reconciling vastly different sampling rates and formats.

More than half of industrial knowledge resides in paper records or experienced operators' tacit knowledge, requiring natural language processing and augmented reality approaches to capture decision-making heuristics. Model interpretability constitutes a fundamental concern where absence of unified validation protocols hinders industrial adoption (Dobbelaere et al., 2021; Romagnoli et al., 2024).

Industry consortia develop benchmark datasets across three dimensions, accuracy under variability, failure detection capability, and compatibility with existing control systems, supporting Safe-and-Sustainable-by-Design initiatives. The computational demands of sophisticated models, while diminishing with hardware advances, present barriers for real-time control requiring millisecond-scale decisions.

Integration with legacy infrastructure demands substantial capital investment and organizational change management, with successful implementation depending critically on collaboration between machine learning experts and domain practitioners. The trajectory of *AI*/*ML* in chemical engineering points toward several transformative directions.

Multi-modal learning frameworks integrating spectroscopic data, process sensor measurements, and mechanistic knowledge will enable more robust predictions across operational regimes. Transfer learning and few-shot learning approaches will extend model applicability to data-scarce applications, particularly for specialized products and emerging processes.

Quantum machine learning, though nascent, promises to revolutionize molecular property prediction and reaction pathway exploration by natively representing quantum mechanical phenomena. Human-*AI* collaboration frameworks will redefine

engineering practice, with *AI* systems augmenting rather than replacing human expertise through explainable recommendations and interactive optimization.

Educational transformation becomes imperative, requiring curricula that integrate data science, machine learning, and traditional chemical engineering principles while emphasizing critical evaluation of model assumptions and limitations. Industry-academia partnerships will accelerate deployment through co-development of validation frameworks, benchmark datasets, and best practice guidelines that balance innovation with operational reliability and regulatory compliance.

9. CONCLUSIONS

The integration of artificial intelligence (*AI*) and machine learning (*ML*) into chemical engineering represents not merely incremental advancement but fundamental reconceptualization of practice spanning molecular design through industrial operations. Contemporary methodologies – from physics-informed neural networks to generative *AI* – demonstrate unprecedented capability in navigating complex multi-scale phenomena while maintaining computational tractability and physical interpretability.

Applications across catalyst discovery, drug development, process control, and sustainability demonstrate transformative impact with quantifiable performance improvements including 71% fewer experimental iterations in reaction optimization, 98% accuracy in predictive maintenance, and substantial reductions in carbon emissions through intelligent process management. Critical challenges persist, demanding concerted efforts in data infrastructure development, interpretability enhancement, and validation protocol standardization.

The path forward requires interdisciplinary collaboration uniting chemical engineering domain expertise with computational intelligence, guided by principles of transparency, reproducibility, and responsible innovation. As the discipline navigates Industry 4.0 transformation toward autonomous, sustainable manufacturing, *AI*/*ML* emerges not as optional enhancement but essential infrastructure enabling the chemical industry to address 21st century imperatives of environmental stewardship, economic competitiveness, and societal benefit.

Success will be measured not only by predictive accuracy but by practical deployment creating measurable value while maintaining safety, sustainability, and operational excellence that define chemical engineering's enduring commitment to serving society. Several overarching observations emerge from a critical synthesis of the literature reviewed here.

First, the field would benefit substantially from a cultural shift toward prospective validation: studies that deploy *AI* models in genuine forward-prediction mode, without access to outcome data during model development, and report performance on truly held-out prospective datasets. The current dominance of retrospective cross-validated benchmarks provides an optimistic but potentially misleading picture of how *AI* systems will perform when confronted with novel process conditions, equipment states, and feedstock compositions not represented in historical data.

Second, the chemical engineering *AI* community needs domain-specific benchmarking standards analogous to those in computer vision (ImageNet) and natural language processing (GLUE). Without agreed benchmark tasks, standardized datasets, and common evaluation protocols, comparing methodologies across groups is unreliable, and apparent progress may partly reflect differences in evaluation rigor rather than genuine advances.

Third, the translation pathway from published algorithm to industrial deployment requires attention to software engineering quality, cybersecurity, operator interface design, and regulatory compliance – dimensions essentially absent from the academic literature but decisive for whether a promising model ever creates real-world value.

Looking forward, the most transformative advances are likely to emerge not from incremental improvements to individual architectures but from principled integration of chemical engineering domain knowledge into learning system design. This means embedding stoichiometric constraints, thermodynamic feasibility bounds, and transport phenomena directly into model structure; developing uncertainty-aware models that reliably distinguish confident in-distribution predictions from speculative extrapolations; and creating continual learning frameworks that adapt to evolving process conditions without catastrophic forgetting of previously validated behavior.

The long-term vision is of autonomous chemical systems combining reinforcement learning optimization, mechanistic model consistency, and explainable *AI* transparency into platforms that genuinely augment engineering practice. Realizing this vision demands that the community resist overstating near-term capabilities and instead invest in the foundational work of rigorous validation, honest benchmarking, and principled uncertainty quantification – the foundations that will ultimately determine whether *AI* becomes reliable infrastructure for chemical manufacturing or remains a collection of promising but fragile demonstrations.

ACKNOWLEDGEMENTS

The manuscript underwent editing by Claude. The author subsequently reviewed the content in its entirety and approved the final manuscript.

REFERENCES

Al Sharah A., Owida H.A., Alnaimat F., Hassan M., Abuowaida S., Alhaj M., Sharadqeh A., 2024. Application of machine learning in chemical engineering: outlook and perspectives. *IAES Int. J. Artif. Intell.*, 13, 619–630. DOI: 10.11591/ijai.v13.i1.pp619-630.

Asrav T., Aydin E., 2023. Physics-informed recurrent neural networks and hyper-parameter optimization for dynamic process systems. *Comput. Chem. Eng.*, 173, 108195. DOI: 10.1016/j.compchemeng.2023.108195.

Benavides-Hernández J., Dumeignil F., 2024. From characterization to discovery: artificial intelligence, machine learning and high-throughput experiments for heterogeneous catalyst design. *ACS Catal.*, 14, 11749–11779. DOI: 10.1021/acscatal.3c06293.

Boskabadi M.R., Murugaiah M., Nielsen T.R., Sivaram A., Sin G., Mansouri S.S., 2025. Virtual sensor for sustainable large-scale process monitoring. *Ind. Eng. Chem. Res.*, 64, 3902–3917. DOI: 10.1021/acs.iecr.4c03342.

Bozal-Ginesta C., Pablo-García S., Choi C., Tarancón A., Aspuru-Guzik A., 2025. Developing machine learning for heterogeneous catalysis with experimental and computational data. *Nat. Rev. Chem.*, 9, 601–616. DOI: 10.1038/s41570-025-00740-4.

Breidung M., Rothe T., Lauff A., Thieme P., Langer J., Günther M., Kuhn H., 2025. Process data-driven machine learning for non-uniformity prediction and virtual metrology in chemical mechanical planarization. *J. Intell. Manuf.* DOI: 10.1007/s10845-025-02753-8.

Burns J.A., Green W.H., 2025. Generalizable, fast, and accurate DeepQSPR with fastprop. *J. Cheminf.*, 17, arXiv:2404.02058v5. DOI: 10.48550/arXiv.2404.02058.

Cao P., Chen J., Chen G., Qi Z., Song Z., 2024. A critical methodological revisit on group-contribution based property prediction of ionic liquids with machine learning. *Chem. Eng. Sci.,* 298, 120395. DOI: 10.1016/j.ces.2024.120395.

Cao X., Zhang Y., Sun Z., Yin H., Feng Y., 2026. Machine learning in polymer science: A new lens for physical and chemical exploration. *Prog. Mater. Sci.*, 156, 101544. DOI: 10.1016/j.pmatsci.2025.101544.

Chinta S., Rengaswamy R., 2019. Machine learning derived quantitative structure property relationship (QSPR) to predict drug solubility in binary solvent systems. *Ind. Eng. Chem. Res.*, 58, 3082–3092. DOI: 10.1021/acs.iecr.8b04584.

Dan A., Liu B., Patil U., Manuraj B.N.M., Gandhi R., Buchel J., Chundawat S.P.S., Guo W., Ramachandran R., 2025. Machine learning model-based design and model predictive control of a bioreactor for the improved production of mammalian cell-based bio-therapeutics. *Control Eng. Pract.*, 156, 106198. DOI: 10.1016/j.conengprac.2024.106198.

Dangayach R., Jeong N., Demirel E., Uzal N., Fung V., Chen Y., 2025. Machine learning-aided inverse design and discovery of novel polymeric materials for membrane separation. *Environ. Sci. Technol.,* 59, 993–1012. DOI: 10.1021/acs.est.4c08298.

De Carfort J.L.N., Pinto T., Krühne U., 2024. An automatic method for generation of CFD-based 3D compartment models: towards real-time mixing simulations. *Bioengineering*, 11, 169. DOI: 10.3390/bioengineering11020169.

Deivendran B., Masampally V.S., Nadimpalli N.R.V., Runkana V., 2025. Virtual metrology for chemical mechanical planarization of semiconductor wafers. *J. Intell. Manuf.*, 36, 1923–1942. DOI: 10.1007/s10845-024-02335-0.

Deng Z., Sarkisov L., 2024. Multi-scale computational design of metal–organic frameworks for carbon capture using machine learning and multi-objective optimization. *Chem. Mater.*, 36, 9806–9821. DOI: 10.1021/acs.chemmater.4c01969.

Dobbelaere M.R., Plehiers P.P., van de Vijver R., Stevens C.V., van Geem K.M., 2021. Machine learning in chemical engineering: strengths, weaknesses, opportunities, and threats. *Engineering*, 7, 1201–1211. DOI: 10.1016/j.eng.2021.03.019.

Eraqi B.A., Khizbullin D., Nagaraja S.S., Sarathy S.M., 2025. Molecular property prediction in the ultra-low data regime. *Commun. Chem.*, 8, 201. DOI: 10.1038/s42004-025-01592-1.

Gao L., Lin J., Wang L., Du L., 2024. Machine learning-assisted design of advanced polymeric materials. *Acc. Mater. Res.*, 5, 571–584. DOI: 10.1021/accountsmr.3c00288.

Gentner N., Carletti M., Kyek A., Susto G.A., Yang Y., 2021. DBAM: Making Virtual Metrology/Soft sensing with time series data scalable through deep learning. *Control Eng. Pract.*, 116, 104914. DOI: 10.1016/j.conengprac.2021.104914.

Georgia Institute of Technology, 2024. Using AI to find the polymers of the future. *ScienceDaily*, 8 November 2024. Available at: https://www.sciencedaily.com/releases/2024/08/240819185140.htm

Goldsmith B.R., Esterhuizen J., Liu J.-X., Bartel C.J., Sutton C., 2018. Machine learning for heterogeneous catalyst design and discovery. *AIChE J.*, 64, 2311–2323. DOI: 10.1002/aic.16198.

He Q.P., Wang J., 2011. Statistics pattern analysis: a new process monitoring framework and its application to semiconductor batch processes. *AIChE J.*, 57, 107–121. DOI: 10.1002/aic.12247.

Heid E., Greenman K.P., Chung Y., Li S.-C., Graff D.E., Vermeire F.H., Wu H., Green W.H., McGill C.J., 2024. Chemprop: a machine learning package for chemical property prediction. *J. Chem. Inf. Model.*, 64, 9–17. DOI: 10.1021/acs.jcim.3c01250.

Hoque A., Surve M., Kalyanakrishnan S., Sunoj R.B., 2024. Reinforcement learning for improving chemical reaction performance. *J. Am. Chem. Soc.*, 146, 28250–28267. DOI: 10.1021/jacs.4c08866.

Kim J., Chang W., Ji H., Joung I.S., 2024. Quantum-informed molecular representation learning enhancing ADMET property prediction. *J. Chem. Inf. Model.*, 64, 5028–5040. DOI: 10.1021/acs.jcim.4c00772.

Lan T., Wang H., An Q., 2024. Enabling high throughput deep reinforcement learning with first principles to investigate catalytic reaction mechanisms. *Nat. Commun.*, 15, 6281. DOI: 10.1038/s41467-024-50531-6.

Lavecchia A., 2025. Explainable artificial intelligence in drug discovery: bridging predictive power and mechanistic insight. *WIREs Comput. Mol. Sci.*, 15, e70049. DOI: 10.1002/wcms.70049.

Lei Q., Li L., Chen H., Wang X., 2023. Emerging directions for carbon capture technologies: a synergy of high-throughput theoretical calculations and machine learning. *Environ. Sci. Technol.*, 57, 17189–17200. DOI: 10.1021/acs.est.3c05305.

Lodh B.K., 2024. The transformative role of Computational Fluid Dynamics (CFD) in chemical engineering. *Open J. Chem.*, 10, 1–3. DOI: 10.17352/ojc.000033.

Manikandan S., Kaviya R.S, Shreeharan D.H, Subbaiya R., Vickram S., Karmegam N., Kim W., Govarthanan M., 2025. Artificial intelligence-driven sustainability: enhancing carbon capture for sustainable development goals – a review. *Sustainable Dev.*, 33, 2004–2029. DOI: 10.1002/sd.3222.

Mathur A., Dave S., Dave S., 2021. Predicting chemical reactions with artificial intelligence and machine learning. In: Purohit S., Singh Jat D., Poonia R., Kumar S., Hiranwal S. (Eds.), *Proceedings of International Conference on Communication and Computational Technologies. Algorithms for Intelligent Systems*. Springer, Singapore, 445–458. DOI: 10.1007/978-981-15-5077-5_36.

Moayedi F., Chandrasekar A., Rasmussen S., Sarna S., Corbett B., Mhaskar P., 2024. Physics-informed neural networks for process systems: handling plant-model mismatch. *Ind. Eng. Chem. Res.*, 63, 13650–13659. DOI: 10.1021/acs.iecr.4c00690.

Mondal S., 2024. Exploring the synergistic potential of artificial intelligence and machine learning in chemistry. *Current Scientia*, 27, 68–85. DOI: 10.31357/vjs.v27i02.7839.

Mou T., Pillai H.S., Wang S., Wan M., Han X., Schweitzer N.M., Che F., Xin H., 2023. Bridging the complexity gap in computational heterogeneous catalysis with machine learning. *Nat. Catal.*, 6, 122–136. DOI: 10.1038/s41929-023-00911-w.

Negri F., Galeazzi A., Gallo F., Manenti F., 2025. Reshaping industrial maintenance with machine learning: fouling control using optimized Gaussian process regression. *Ind. Eng. Chem. Res.*, 64, 6633–6654. DOI: 10.1021/acs.iecr.4c04550.

Niazi S.K., 2025. Artificial intelligence in small-molecule drug discovery: a critical review of methods, applications, and real-world outcomes. *Pharmaceuticals*, 18, 1271. DOI: 10.3390/ph18091271.

Niu Z., Xiao X., Wu W., Cai Q., Jiang Y., Jin W., Wang M., Yang G., Kong L., Jin X., Yang G., Chen H., 2024. PharmaBench: enhancing ADMET benchmarks with large language models. *Sci. Data*, 11, 985. DOI: 10.1038/s41597-024-03793-0.

Omidvar N., Wang S.-H., Huang Y., Pillai H.S., Athawale A., Wang S., Achenie L.E.K., Xin H., 2024. Explainable AI for optimizing oxygen reduction on Pt monolayer core–shell catalysts. *Electrochem. Sci. Adv.*, 4, e202300028. DOI: 10.1002/elsa.202300028.

Omranpour A., Elsner J., Lausch K.N., Behler J., 2025. Machine learning potentials for heterogeneous catalysis. *ACS Catal.*, 15, 1616–1634. DOI: 10.1021/acscatal.4c06717.

Oviedo F., Ferres J.L., Buonassisi T., Butler K.T., 2022. Interpretable and explainable machine learning for materials science and chemistry. *Acc. Mater. Res.*, 3, 597–607. DOI: 10.1021/accountsmr.1c00244.

Patel R., Bhartiya S., Gudi R., 2024. Model predictive control using physics-informed neural networks for process systems. *IFAC-PapersOnLine*, 58, 775–780. DOI: 10.1016/j.ifacol.2024.08.431.

Pathan I., Raza A., Sahu A., Joshi M., Sahu Y., Patil Y., Raza M.A., 2025. Revolutionizing pharmacology: AI-powered approaches in molecular modeling and ADMET prediction. *Med. Drug Discovery*, 28, 100223. DOI: 10.1016/j.medidd.2025.100223.

Rahman A.U., Han X., Jia X., 2024. Enhanced data-driven virtual metrology on chemical mechanical planarization process using dual linear Kalman filter. *Proceedings of the ASME 2024 19th International Manufacturing Science and Engineering Conference*. Knoxville, Tennessee, USA. June 17–21, 2024. V002T05A009. ASME, 1–8. DOI: 10.1115/MSEC2024-125308.

Romagnoli J.A., Briceño-Mena L., Manee V., 2024. *AI in Chemical Engineering: Unlocking the power within data*. CRC Press. DOI: 10.1201/9781003455905.

Sanjuan Gómez A., 2021. *Fault diagnosis and fault-tolerant control of nonlinear dynamic systems using artificial intelligence techniques*. Doctoral dissertation, Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya (UPC). DOI: 10.5821/dissertation-2117-420260.

Smart E.E., Olayiwola D.E., Okwor U.D., Babatunde J., 2025. Revolutionizing carbon capture efficiency: a comprehensive review of AI-driven optimization strategies for CO2 capture and storage processes. *Int. J. Sci. Technol. Res. Arch.*, 9, 18–43. DOI: 10.53771/ijstra.2025.9.1.0043.

Swanson K., Walther P., Leitz J., Mukherjee S., Wu J.C., Shivnaraine R.V., Zou J., 2024. ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics*, 40, btae416. DOI: 10.1093/bioinformatics/btae416.

Terstappen G.C., Reggiani A., 2001. *In silico* research in drug discovery. *Trends Pharmacol. Sci.*, 22, 23–26. DOI: 10.1016/S0165-6147(00)01584-4.

Velioglu M., Zhai S., Rupprecht S., Mitsos A., Jupke A., Dahmen M., 2025. Physics-informed neural networks for dynamic process operations with limited physical knowledge and data. *Comput. Chem. Eng.*, 192, 108899. DOI: 10.1016/j.compchemeng.2024.108899.

Venkataraman M., Rao G.C., Madavareddi J.K., Maddi S.R., 2025. Leveraging machine learning models in evaluating ADMET properties for drug discovery and development. *ADMET and DMPK*, 13, 2772. DOI: 10.5599/admet.2772.

Vora L.K., Gholap A.D., Jetha K., Thakur R.R.S., Solanki H.K., Chavda V.P., 2023. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics*, 15, 1916. DOI: 10.3390/pharmaceutics15071916.

Wang C.-C., 2024. T^2-LSTM-based AI system for early detection of motor failure in chemical plants. *Mathematics*, 12, 2652. DOI: 10.3390/math12172652.

Wang R., Fu T., Yang Y.-J., Song X., Wang X.-L., Wang Y.-Z., 2024a. Scientific discovery framework accelerating advanced polymeric materials design. *Research*, 7, 0406. DOI: 10.34133/research.0406.

Wang W., Zhang H., Wang Y., Tian Y., Wu Z., 2024b. Fast explicit machine learning-based model predictive control of nonlinear processes using input convex neural networks. *Ind. Eng. Chem. Res.*, 63, 17279–17293. DOI: 10.1021/acs.iecr.4c02257.

Wellawatte G.P., Schwaller P., 2025. Human interpretable structure-property relationships in chemistry using explainable machine learning and large language models. *Commun. Chem.*, 8, 11. DOI: 10.1038/s42004-024-01393-y.

Wenzel J., Matter H., Schmidt F., 2019. Predictive multitask deep neural network models for ADME-Tox properties: learning from large data sets. *J. Chem. Inf. Model.*, 59, 1253–1268. DOI: 10.1021/acs.jcim.8b00785.

Wieder O., Kohlbacher S., Kuenemann M., Garon A., Ducrot P., Seidel T., Langer T., 2020. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technol.*, 37, 1–12. DOI: 10.1016/j.ddtec.2020.11.009.

Wu M., Di Caprio U., Vermeire F., Hellinckx P., Braeken L., Waldherr S., Leblebici M.E., 2023. An artificial intelligence course for chemical engineers. *Educ. Chem. Eng.*, 45, 141–150. DOI: 10.1016/J.ECE.2023.09.004.

Wu S., Kondo Y., Kakimoto M.-A., Yang B., Yamada H., Kuwajima I., Lambard G., Hongo K., Xu Y., Shiomi J., Schick C., Morikawa J., Yoshida R., 2019. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput. Mater.*, 5, 66. DOI: 10.1038/s41524-019-0203-2.

Wu Z., Li M., He C., Zhang B., Ren J., Yu H., Chen Q., 2024. Physics-informed learning of chemical reactor systems using decoupling-coupling training framework. *AIChE J.*, 70, e18436. DOI: 10.1002/aic.18436.

Xiao T., Wu Z., Christofides P.D., Armaou A., Ni D., 2021. Recurrent neural-network-based model predictive control of a plasma etch process. *Ind. Eng. Chem. Res.*, 61, 638–652. DOI: 10.1021/acs.iecr.1c04251.

Xin H., Mou T., Pillai H.S., Wang S.-H., Huang Y., 2023. Interpretable machine learning for catalytic materials design toward sustainability. *Acc. Mater. Res.*, 5, 22–34. DOI: 10.1021/accountsmr.3c00131.

Yan Y., Borhani T.N., Clough P.T., 2020. Chapter 14: Machine Learning Applications in Chemical Engineering, In: Cartwright H.M. (Ed.), *Machine learning in chemistry: The impact of artificial intelligence*. The Royal Society of Chemistry, 340–371. DOI: 10.1039/9781839160233-00340.

Yang K., Swanson K., Jin W., Coley C., Eiden P., Gao H., Guzman-Perez A., Hopper T., Kelley B., Mathea M., Palmer A., 2019. Analyzing learned molecular representations for property prediction. *J. Chem. Inf. Model.*, 59, 3370–3388. DOI: 10.1021/acs.jcim.9b00237.

Yue T, He J, Li Y. Machine-Learning-assisted molecular design of innovative polymers. 2025. *Acc. Mater. Res.*, 6, 1033–1045. DOI: 10.1021/accountsmr.5c00151.

Zhang S., Yan Z., Huang Y., Liu L., He D., Wang W., Fang X., Zhang X., Wang F., Wu H., Wang H., 2022. HelixADMET: a robust and endpoint extensible ADMET system incorporating self-supervised knowledge transfer. *Bioinformatics*, 38, 3444–3453. DOI: 10.1093/bioinformatics/btac342.

Zhou Z., Li X., Zare R.N., 2017. Optimizing chemical reactions with deep reinforcement learning. *ACS Cent. Sci.*, 3, 1337–1344. DOI: 10.1021/acscentsci.7b00492.