Preprints.org

Review

Not peer-reviewed version

Bridging the Reality Gap: Machine Learning, Live Analytics, and the New Practice of Predictive Chemistry

Samrat Chakraborty *

Posted Date: 15 October 2025

doi: 10.20944/preprints202510.1185.v1

Keywords: predictive chemistry; reality gap (theory–experiment); hybrid QM/ML; Δ-learning error correction; physics-informed machine learning; uncertainty quantification; calibrated confidence; uncertainty propagation into decisions; transfer learning; domain adaptation; OOD robustness; active learning; Bayesian optimization; closed-loop experimentation; conditional autonomy; human-in-the-loop robotics

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

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

Review

Bridging the Reality Gap: Machine Learning, Live Analytics, and the New Practice of Predictive Chemistry

Samrat Chakraborty

Indian Association for the Cultivation of Science (IACS), Kolkata, India; scssc3357@iacs.res.in

Abstract

Every chemist knows the small heartbreak: the calculation looks beautiful; the flask does not. This paper takes that feeling seriously and names it—the reality gap—and then shows how to cross it on purpose. Our thesis is straightforward: predictive chemistry emerges when we let theory and experiment argue in public, with machine learning acting as the translator that keeps the debate honest. We first map where neat first-principles wobble in the wild: bonds stretching and breaking, surfaces choosing a pathway, solvents shifting free energies just enough to matter, and spin states reorganizing the landscape. We then show how to correct those edges without discarding physics: hybrid QM/ML methods that learn systematic errors, uncertainty that travels with every prediction so we know when to trust and when to measure, and chemistry-aware transfer learning so models trained on idealized inputs remain useful on real instruments. The loop closes when models talk to tools and tools talk back. Process-analytical technologies feed real-time signals to Bayesian optimization, multi-objective workflows make trade-offs visible (yield, selectivity, cost, greenness), and autonomy becomes conditional by design—robots execute, chemists steer. We focus on validation that survives deployment, not convenience: splits that reflect how chemistry varies in practice, calibrated confidence, and structured logging that treats failures as first-class data. Finally, we detail what this buys in real laboratories: faster cycles, reproducible and information-rich datasets, greener routes—and decisions made with eyes open. The message is practical and hopeful. Keep the physics where it is strong. Teach it where it is stubbornly wrong. Carry uncertainty forward. Let instruments help decide where to look next. Do that, and the calculation and the flask still won't always agree—but they will disagree productively, more often, and for reasons we can understand. That is predictive chemistry in practice.

Keywords: predictive chemistry; reality gap (theory–experiment); hybrid QM/ML; Δ-learning error correction; physics-informed machine learning; uncertainty quantification; calibrated confidence; uncertainty propagation into decisions; transfer learning; domain adaptation; OOD robustness; active learning; Bayesian optimization; closed-loop experimentation; conditional autonomy; human-in-the-loop robotics

---

Introduction

Every chemist knows the same small heartbreak: the calculation looks beautiful; the flask does not. On the screen, the potential-energy surface is smooth and the barrier is friendly; at the bench, solvent, surface, spin state, or a trace impurity rewrites the plot. That mismatch isn't just a vibe—it's measurable. "Chemical accuracy" hovers around 1 kcal/mol for bond-making and bond-breaking, yet the places we care about most still trip our best tools: static correlation, strongly interacting media, and open-shell transition-metal chemistry. For a simple C–C bond dissociation, lower-rung functionals can miss by 30–50 kcal/mol, and even sophisticated methods wobble unless they learn to correct those deficiencies [1]. Move from gas phase to solution and you inherit concentration-dependent free-energy shifts of several kcal/mol; implicit solvation alone often isn't

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

enough, and cluster–continuum treatments are needed to catch short-range structure and hydrogen bonding [2] None of this is a dunk on quantum chemistry—DFT remains the workhorse across biochemistry, catalysis, and materials [3] It's just an honest map of where approximation meets the world.

Reality gets messier when models leave the sandbox of curated benchmarks and meet live data streams. Domain shift, hidden stratification, and plain old leakage can make paper-perfect models stumble in deployment [4]. Meanwhile, the space we must navigate is astronomical: in retrosynthesis, more than 10,000 plausible disconnections can appear at a single step, and the search tree explodes with every additional move [5]. We rarely have the luxury of exhaustive trial-and-error to rescue theory with brute force.

What's changed—quietly but decisively—is that machine learning has matured into a translator between theory and experiment. Not a replacement for physics, but an adaptive layer that learns systematic errors, quantifies uncertainty, and closes the loop with data. On the physics side, hybrid strategies such as Δ-ML, ML-enhanced exchange–correlation functionals, and learned double-hybrid schemes reduce stubborn errors without abandoning first-principles foundations. Skala learns directly from high-level reference data to reach chemical-accuracy atomization energies while retaining the computational efficiency of semi-local DFT [6]. R-xDH7 explicitly targets static and dynamic correlation together, shrinking difficult bond-dissociation errors to the right scale for chemistry [1]. On the decision side, uncertainty quantification turns predictions into testable statistical hypotheses—telling us when to trust a calculation and when to measure instead [7,8]. And when ab initio data are plentiful but imperfect while experimental data are sparse but decisive, transfer learning and domain adaptation knit the two so that models trained on "idealized" inputs remain useful in the lab [9].

The moment these pieces connect to instruments, they stop being abstractions. Process-analytical technologies feed real-time spectroscopic signals into Bayesian optimization loops that make uncertainty a feature, not a bug; multi-objective formulations balance yield, selectivity, cost, and greenness instead of chasing a single number [10]. This is the engine room of today's self-driving laboratories: robots and flow platforms that can run hundreds of experiments while you sleep—and more importantly, learn from each one to steer the next [11,12]. It works in practice. Learning from "failed" hydrothermal syntheses let an ML model propose crystallization conditions for templated vanadium selenites with an 89% experimental success rate—outperforming human intuition precisely because it mined the dark data we usually discard [13]. In materials growth, the CARCO workflow combined language models, automation, and data-driven optimization to rapidly home in on catalysts and process windows for high-density aligned CNT arrays, compressing months of trial-and-error into weeks of guided exploration [14]. In organic synthesis, real-time in-line NMR has already been used to optimize catalytic reactions on the fly, not after the fact [15].

This perspective is about making that bridge explicit. First, we'll put numbers and mechanisms to the gap—where and why first-principles miss, from static correlation and surfaces to solvation and spin [1,2]. Then we'll show how machine learning acts as a universal translator: physics-informed and hybrid QM/ML models that correct systematic errors; uncertainty frameworks that turn outputs into hypotheses; and transfer learning that carries models from simulation to experiment without wishful thinking [6,7].Finally, we'll follow the loop into the lab—Bayesian, real-time, multi-objective optimization coupled to PAT and automation—anchoring the story in case studies where the loop held up under experimental pressure [10,11]. The goal isn't to pick sides. It's to show, with evidence, how computation and experiment finally inform one another in a self-correcting cycle—and what it will take to make predictive chemistry routine rather than exceptional.

Machine Learning as the Universal Translator

We don't need to replace physics—we need to teach it where it goes wrong, quantify that miss honestly, and then use data to close the loop. That is what modern machine learning is good at in chemistry: it acts as a translator between neat first-principles predictions and messy experimental

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

reality, learning systematic errors, expressing uncertainty, and steering us toward the most informative next measurement.

2.1. Physics-Informed Models: Build the Rules into the Learner

Chemistry rewards models that respect its laws. Physics-informed approaches embed those laws directly into training, so models don't just fit—they behave. In practice, this looks like hard constraints that enforce thermodynamic consistency and symmetry, and loss functions that include the residuals of governing equations. The point is simple: if the learning algorithm knows the rules, it needs less data, extrapolates more gracefully, and avoids predictions that might look clever but violate first principles [16]. In parallel, symmetry-aware graph neural networks and physics-embedded neural architectures improve sample efficiency when data are scarce and the stakes for consistency are high [9,16].

2.2. Hybrid QM/ML: Correct What Physics Misses, Don't Replace It

The fastest, most transparent path from a beautiful calculation to a usable prediction is often a corrective layer, not a reinvention. Hybrid QM/ML methods do exactly that. Skala learns exchange–correlation directly from high-level reference data and reaches chemical-accuracy atomization energies while retaining the computational efficiency typical of semi-local DFT [6]. R-xDH7 combines machine learning with a renormalized double-hybrid formulation to treat static and dynamic correlation together, shrinking difficult bond-dissociation errors into the ~1 kcal/mol neighborhood chemistry demands [1]. Closely related Δ-learning strategies correct the systematic gap between a cheap baseline (for example, HF or semi-empirical methods) and a trusted reference, while symmetry-aware models like OrbNet elevate semi-empirical electronic structure toward DFT-level fidelity with far lower cost [9]. The philosophy is consistent: keep the physics where it is strong; let the data fix what it chronically misses.

2.3. Transfer Learning from Simulation to Experiment

Ab initio datasets are abundant but idealized; experimental datasets are definitive but sparse. Bridging them requires transfer learning and domain adaptation that are aware of chemistry, not just statistics. Chemistry-informed domain transformation maps quantities learned in simulation to their experimental counterparts using known physical relationships and statistical ensembles, so models trained on DFT can be fine-tuned to reality with minimal lab data [17]. Mult fidelity learning goes a step further by combining cheap, noisy computational data with expensive, accurate references to reach practical accuracy at a fraction of the cost [9]. The message is practical: design the Sim-to-Real pipeline so the model remains useful on your instrument, not just impressive on synthetic benchmarks.

2.4. Uncertainty as a First-Class Signal

A prediction without its uncertainty is a guess. In chemistry, calibrated uncertainty is a design variable—it tells you when to trust a calculation, and when to measure instead. By propagating uncertainty in established reactivity scales, we can convert point predictions into testable statistical hypotheses and design experiments that are maximally discriminative [7]. Calibrated approaches, from ensembles to Bayesian layers, make coverage explicit so the next experiment is chosen where it improves both outcome and understanding [8]. Uncertainty is not a nuisance here; it is the compass that points to the next best experiment.

2.5. Closing the Loop with Instruments: PAT and Bayesian Optimization

The translator becomes a pilot when it connects to instruments. Process-analytical technologies—real-time NMR, IR, MS—feed Bayesian optimization loops that treat uncertainty as an asset rather than a flaw. Instead of hoping a preplanned grid happens to land on the sweet spot, the

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

algorithm targets regions where the model is uncertain and an experiment would change our belief the most [10] This isn't theoretical. Real-time in-line NMR has already been used to optimize catalytic organic reactions on the fly, folding stereochemical and multinuclear readouts into live decisions [15]. And because chemistry rarely has a single objective, modern workflows make trade-offs explicit: multi-objective optimization balances yield and selectivity with cost, safety, and greenness, and the Pareto front shows what is possible—and what must be sacrificed—before any decision is locked in [10]

2.6. Case Studies That Prove the Bridge

The most convincing argument is a result that holds up in glass. Learning from "failed" hydrothermal syntheses allowed a recommender to propose crystallization conditions for templated vanadium selenites with an 89% experimental success rate, outperforming human intuition precisely because it learned from the outcomes we usually ignore [13]. In materials growth, the CARCO platform combined language models, automation, and data-driven optimization to rapidly home in on catalysts and process windows for high-density aligned CNT arrays, compressing what would have been months of trial-and-error into weeks of guided exploration [14]. And in organic synthesis, real-time in-line NMR closed the loop for catalytic reactions, replacing after-the-fact autopsies with live control [15]. These aren't just elegant workflows; they are concrete proof that the translation holds under experimental pressure.

Transition

With the translator in place—physics corrected where it falters, uncertainty carried honestly, and instruments listening—we can finally judge the field by what matters: how often predictions work at the bench, how validation should be done so results travel, and what all this buys us in time, reproducibility, and greener choices. That is the focus of the next section.

Closing the Loop: Real-Time Optimization and Autonomous Platforms

When models talk to instruments—and instruments talk back—prediction stops being a static claim and becomes a living process. This is the point where uncertainty guides the next experiment, where trade-offs are made explicit rather than hidden, and where robots shoulder the repetitive work so chemists can steer.

3.1. Real-Time Analytics with Bayesian Optimization

The core idea is elegantly simple: let the experiment stream its own data while it runs, and use that stream to decide, in the moment, what to try next. Process-analytical technologies—such as in-line NMR, IR, or MS—close the gap between measurement and decision, and Bayesian optimization turns model uncertainty into a feature rather than a flaw. Instead of laying down a fixed grid and hoping to land on a sweet spot, the optimizer probes where the model is least certain and where a new data point would change our belief the most [10]. This isn't just an attractive diagram. Real-time in-line NMR has already been used to steer catalytic organic reactions on the fly, folding stereochemical and multinuclear readouts into live decisions and replacing after-the-fact autopsies with genuine control [15].

3.2. Multi-Objective Workflows: Making Trade-Offs Visible

Chemistry rarely optimizes a single number. Yield and selectivity matter, but so do cost, safety, and environmental footprint. Modern workflows move these goals out of the footnotes and onto the same axis as performance, using multi-objective optimization to draw a Pareto front that shows what is possible—and what must be sacrificed—before any decision is locked in [10]. The effect is tangible at the route level. In the total synthesis of a helicase-primase inhibitor, computer-aided retrosynthesis paired with human guidance increased overall yield from 8% to 26%, improved greenness scores,

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

and with further refinement reached 35% while dramatically reducing building-block costs [18]. The practical lesson is uncomplicated: if you ask for multiple objectives, you can actually get them.

3.3. Conditional Autonomy and Human-in-the-Loop Control

Autonomy worth trusting is never autonomy without oversight. The most effective self-driving laboratory platforms are explicitly "conditionally autonomous": they run independently under known conditions, surface anomalies, and invite human guidance when something looks off [11]. In practice, that looks like workflow mutability—reordering steps or swapping sensors when data drift is detected—and clear uncertainty thresholds that trigger a pause for review. The chemist's role shifts from manual executor to strategist and arbiter; the work becomes less about pipetting and more about decision-making.

3.4. Throughput, Reproducibility, and Why It Matters

Numbers focus attention. At the University of Liverpool, a mobile robotic chemist autonomously performed 688 experiments in eight days and identified photocatalysts that were six times more active than prior baselines—work that would have taken a human team months [12]. Beyond speed, digital orchestration turns reproducibility into the default: protocols are encoded, measurements are time-stamped, and every decision is auditable [11]. That is not bureaucracy; it is how we build datasets truly fit for machine learning—internally consistent, rich in context, and inclusive of the negative results we used to throw away.

3.5. Case Studies: Proof That the Loop Holds

The most persuasive argument remains a result that survives contact with glassware. Learning from "failed" hydrothermal syntheses allowed a recommender to propose crystallization conditions for templated vanadium selenites with an 89% experimental success rate, outperforming human intuition by leaning on the outcomes we usually ignore [13]. In materials growth, the CARCO platform combined language models, automation, and data-driven optimization to rapidly home in on catalysts and process windows for high-density aligned CNT arrays, compressing what would have been months of trial-and-error into weeks of guided exploration [14]. And in organic synthesis, in-line NMR closed the loop for catalytic reactions, turning reaction monitoring into a lever for real-time control [15]. Together, these examples show the same pattern: uncertainty-aware models, live analytics, and adaptive execution make "promising in silico" become "works in glass" more often—and much faster.

Transition

With the loop closed—models listening to instruments, instruments informing models, and chemists steering—we can now judge the field by the only yardsticks that matter: validation that travels from paper to bench, and concrete gains in time, cost, reproducibility, and greener choices. The next section focuses on how to measure that impact fairly and what robust deployment looks like in real laboratories.

Validation and Real-World Performance

A model only earns its keep when its suggestions work in glass. In this section, we stay with the bench: how often machine-learning-assisted decisions succeed in real experiments, where they falter, how validation should be designed so results travel, and how performance changes when we move from tidy model systems to unruly, real-world substrates.

4.1. Experimental Success: What Survives Contact with the Bench

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

Some of the clearest evidence that data-driven methods bridge theory and practice comes from campaigns that test predictions directly. A landmark example is the "learning from failure" study on templated vanadium selenites. By training on both successful and "dark" (failed) hydrothermal syntheses, the recommender proposed new crystallization conditions that succeeded in 89% of experimental tests—outperforming human intuition precisely because it mined the outcomes we usually discard [13].Reaction-site prediction tells a similar story in organic synthesis: for nucleophilic aromatic substitution (SNAr), a hybrid approach that blends mechanistic modeling with learned corrections delivered mid-80s top-1 accuracies for both regioselectivity and chemoselectivity, significantly improving over rankings based solely on computed barriers [19]. At scale, condition recommenders trained on roughly ten million reactions from Reaxys can place a near-correct "chemical context" (catalyst plus at least one solvent/reagent) within the top-10 suggestions for the majority of held-out cases; individual components often land in the 80–90% top-10 range, and temperature estimates fall within ±20 °C for most examples—sharpening when the broader context is correct [20].

4.2. Academic "Wins" Versus Industrial Reality

How we validate matters as much as what we validate. Models that look pristine under random cross-validation can stumble as soon as they meet genuinely new batches, instruments, or substrates—classic domain shift and hidden stratification. In materials case studies, near-perfect R [2] scores under k-fold fell to roughly the 0.5–0.6 range when "leave-one-lot-out" protocols were used to mimic deployment, revealing exactly how fragile over-optimistic splits can be [4]. Chemistry faces the same risks: if training and test sets share scaffolds, lab signatures, or text-mined artifacts, we will overestimate how predictive our predictions really are. This is why industrial adoption often looks pragmatic rather than theatrical. At AstraZeneca, for example, AiZynthFinder is used at scale to process thousands of compounds per month, with median search times of minutes and prompting features that let chemists constrain routes by "freezing" or "breaking" specific bonds—an explicitly human-guided, fitness-for-purpose workflow rather than a quest for headline accuracy [21].

4.3. Where Predictions Break: Recurring Failure Modes

Patterns repeat across domains. Data leakage and split bias inflate apparent generalization whenever near-duplicates straddle train and test. Literature-trained condition recommenders tend to reproduce popular contexts rather than discover optimal ones for a new substrate, which makes them excellent starting points but poor finish lines [20]. The absence of negative data—failed reactions and low-yield outcomes—blurs decision boundaries and breeds overconfidence; ultrahigh-throughput experimentation was developed in part to correct this by capturing successes and failures in structured, machine-readable form [22]. Instrument drift and subtle protocol changes can quietly shift the data distribution under a trained model, degrading predictions even when the underlying chemistry hasn't changed [4]. None of these issues are exotic; they are what deployment looks like.

4.4. From Model Compounds to Complicated Substrates

Generalization depends on both chemistry and data regime. When only a few dozen examples exist in a reaction family, mechanistic or physics-guided models typically beat purely data-driven ones; at intermediate sample sizes, hybrids (mechanism plus ML) tend to be most accurate; as datasets grow into the hundreds, physical-organic QSRR and then structure-based models become increasingly competitive [19]. Transfer learning provides a practical bridge. By pre-training on abundant ab initio data and fine-tuning on sparse experimental measurements—using chemistry-informed domain transformations or multifidly pipelines—we can keep models useful on instruments rather than only on idealized simulations [9,17]. The lesson is straightforward: match the modelling strategy to the data regime you actually have, and design your Sim-to-Real pipeline so it survives contact with your lab, not just your benchmark.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

4.5. What Good Validation Looks Like (So Results Travel)

Robust deployments share a few traits. First, they validate the way they will be used: scaffold-wise or lot-wise splits rather than random ones; explicit inclusion of instrument and batch variation; and tests that probe the conditions most likely to shift in practice [4]. Second, they report calibrated uncertainty so chemists know when to trust a prediction—and when to measure instead; turning predictions into testable statistical hypotheses makes experimental design more efficient and conclusions more honest [7,8]. Third, they treat failures as first-class data. Ultrahigh-throughput experiments and structured logging of negative outcomes sharpen decision boundaries and make déjà vu projects less likely [13,22]. Finally, they separate "popular" from "optimal" by re-optimizing promising ML-suggested contexts on new substrates rather than assuming historical recipes transfer for free [20]. Validation isn't window dressing here; it is the difference between an attractive figure and a tool that changes how we work.

Transition

With validation designed to survive deployment, we can ask the only question that matters outside the slide deck: what does this buy us? The next section turns to time-to-insight, reproducibility, and greener decisions—where data-rich, uncertainty-aware, closed-loop workflows have already shifted the economics and the culture of discovery [10,18].

Economic and Practical Impact

Predictive chemistry has to pay its way. The promise is not clever figures; it is faster cycles, cleaner data, greener routes, and decisions made with eyes open. Here we stay concrete: what changes when machine learning, high-throughput analytics, and automation leave the slide deck and run in real labs.

5.1. Time-to-Insight and Throughput

Analytical speed sets the rhythm for discovery. Desorption electrospray ionization mass spectrometry has pushed reaction-array readouts from hours to minutes, enabling order-of-magnitude faster screening and analysis at microliter–nanoliter scales. In practice, DESI-MS has delivered reaction screening at rates approaching ten thousand reactions per hour, with rapid characterization that collapses the analysis bottleneck [23–25]. Acoustic Mist Ionization scales that paradigm further, enabling contactless, ultrahigh-throughput sampling on the order of one hundred thousand samples per day from a single mass spectrometer [26].

On the execution side, automation compresses months of manual iteration into days of orchestrated work. At the University of Liverpool, a mobile robotic chemist autonomously executed 688 experiments in eight days and uncovered photocatalysts six times more active than prior baselines—a workload that would have taken a human team months [11,12]. In flow, feedback is immediate: automated stopped-flow libraries and real-time analytics enable hundreds to thousands of experiments per week while consuming a fraction of the reagents and solvents used in batch campaigns [15,27].

5.2. Cost, ROI, and How to Scale Wisely

The quickest returns from digital and automated workflows are not just per-experiment savings; they are strategic time reclaimed by avoiding false paths and repeating inconsistent work. Organizations that invest in orchestration—well-specified digital protocols, structured data, automated analysis—report cumulative savings from reduced waste, higher instrument utilization, and fewer non-value-adding iterations [28]. Scale matters: "numbered-up" microreactor arrays with extensive control hardware can deliver outsized benefits for high-value programs or where they unlock dramatic yield or safety gains, but modular, loosely integrated setups allow teams to see returns earlier and de-risk expansion [29].Macro-level trends point in the same direction. Analyses of

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

generative AI as per McKinsey company suggest a 0.1–0.6% annual uplift in labor productivity on its own and 0.5–3.4% when combined with broader automation—an indication of what integrated code–data–instrument ecosystems can deliver when deployed at scale [30].

5.3. Reproducibility and Data Quality (Including the "Dark" Data We Used to Throw Away)

Digitally encoded protocols, time-stamped events, and auditable decisions turn reproducibility into the default rather than the exception. That is not bureaucracy—it is how we generate datasets that are genuinely fit for machine learning [11]. Ultrahigh-throughput experimentation complements this by capturing both successes and failures in machine-readable form, sharpening decision boundaries where purely literature-mined datasets are biased toward "greatest hits" [22]. The payoff is visible in practice. By training on failed hydrothermal syntheses as well as successes, a recommender suggested crystallization conditions for templated vanadium selenites that succeeded in 89% of experimental tests—outperforming expert intuition by learning from the outcomes we usually hide [13]. Negative results are not noise; they are the contours of the map.

5.4. Sustainability and Greener Routes

When cost and environmental impact are treated as first-class objectives alongside yield and selectivity, synthesis planning changes. In the total synthesis of a helicase-primase inhibitor API, computer-aided retrosynthesis paired with human guidance increased overall yield from 8% to 26%, improved greenness metrics, and with further human-guided refinement reached 35% while dramatically reducing building-block costs [18]. The principle generalizes: define the Pareto front (productivity, selectivity, cost, greenness), then choose with trade-offs visible rather than hidden in a single metric [10].

5.5. Sector-Specific Impacts

In pharmaceuticals, flow-based feedback and high-throughput experimentation have become integral to route scouting, catalyst discovery, and scale-up, particularly for cross-couplings, photoredox catalysis, and asymmetric transformations [29,31]. In materials and polymers, machine learning drives property prediction and discovery loops while automation closes the gap between process parameters and performance; the CARCO platform's rapid optimization of catalysts and growth windows for high-density, aligned carbon nanotube arrays is a striking case [14]. In fine chemicals, continuous microreactor technologies often deliver decisive gains in yield and safety, making them attractive when economics and risk profiles align [29].

Transition

If Section 2 established how machine learning translates physics into practice—and Section 3 showed how instruments and algorithms learn from each other in real time—this section answers the only question that matters outside the slide deck: what does it buy us? Faster cycles, structured and reproducible data, greener routes, and better choices are not aspirational; they are already here in the right settings. In the final section, we look ahead at the infrastructure and guardrails that will make predictive chemistry routine rather than exceptional: digital twins for live model updating, quantum acceleration where it matters, immersive interfaces that improve oversight and training, and the ethical scaffolding that keeps autonomy safe and accountable.

Future Frontiers and Emerging Challenges

The bridge between computation and experiment is now real: physics where it is strong, learning where it is weak, and instruments listening in real time. What comes next are the infrastructures and guardrails that make predictive chemistry routine—digital twins that keep models synchronized with reality, quantum acceleration where it matters, immersive interfaces that improve oversight and training, and ethical scaffolding that keeps autonomy safe and accountable.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

6.1. Next-Generation Integration Technologies

6.1.1. Quantum Computing as an Accelerator of First-Principles

Quantum algorithms are being developed precisely for the hard corners of electronic structure—where static correlation, multireference character, and delicate energy differences strain classical methods. Approaches such as qubit coupled-cluster ansätze, variational quantum eigensolvers, and related multireference schemes aim to compute ground and excited states with higher fidelity, especially in strongly correlated regimes. The practical impact for integration is twofold. First, better potential-energy surfaces and properties feed downstream machine-learning correctors with cleaner targets, shrinking the "reality gap" at its source. Second, hybrid classical–quantum workflows—where a quantum kernel handles the intractable subproblem and classical layers learn residual corrections—align naturally with the corrective philosophy outlined here [32]. The promise is substantial, but the mindset remains the same: deploy quantum acceleration where it measurably improves the decisions your lab must make, and keep validation grounded in experiment [4].

6.1.2. Digital Twins: Keeping Models Honest in Real Time

A digital twin is a live, data-driven mirror of a physical process. In chemistry, that means a virtual reactor or workflow that ingests real-time measurements, updates kinetic or statistical models on the fly, and predicts the consequences of the next change before you make it [33]. The mechanics are straightforward: data-centric engineering synchronizes simulation, machine learning, and statistics so that the twin evolves as the experiment evolves [33]. In practice, this enables automated kinetic model discovery while experiments are still running, and it pairs naturally with active learning—each new measurement is chosen to be maximally informative, reducing the experimental burden rather than just increasing the dataset [34].

6.1.3. AR/VR for Oversight, Training, and "Touching" Molecules

Immersive interfaces are not a gimmick when the systems are complex and the stakes are high. Augmented reality can render 3D molecular structures and reaction pathways in situ, overlaying procedural cues and safety information directly onto the lab environment) [35]. Virtual reality goes deeper—letting chemists explore protein–ligand complexes, reaction coordinates, or process layouts in a space where scale and geometry are intuitive. Paired with computer vision and autonomy, these tools improve human oversight, speed training, and give operators a more intuitive grasp of what the automation is doing and why [36]

6.1.4. Toward Fully Autonomous Discovery (With a Human on the Loop)

Self-driving laboratories already integrate hypothesis generation, design, execution, and analysis; the path forward is higher fidelity and broader scope, not magic. Hardware and software stacks are emerging for universal chemical synthesis and exploration, from standardized languages like XDL for protocol portability to orchestration layers that coordinate robotics, analytics, and optimization [36,37]. On the decision side, cost-informed Bayesian optimization, self-optimization algorithms, and multi-objective search formalize the trade-offs chemists actually care about—yield, selectivity, cost, and greenness—rather than chasing a single metric [38–40]. The destination is not "no humans." It is conditional autonomy with clear escalation triggers—systems that propose and execute most steps confidently, surface anomalies with calibrated uncertainty, and invite human judgment when it matters [36].

6.2. Ethical and Societal Implications

6.2.1. Skills and Roles: A Shift, Not a Replacement

Automation changes what chemists do; it does not make chemists optional. The immediate need is hybrid literacy: experimentalists who can read and shape digital workflows, and computational

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

scientists who understand lab realities [9,38]. Many academic programs still under-teach modern optimization (DoE, HTE) and orchestration; industry practice already expects it [38]. The payoff is that humans move up the stack—problem selection, hypothesis framing, anomaly adjudication, and strategic decision-making—while robots handle repetition [39].

6.2.2. Safety and Security: Design for the Worst Day, Not the Best

Some reactions—palladium-catalyzed cross-couplings among them—can exhibit significant exotherms, with runaways possible if heat is not actively managed. Autonomous systems must be designed around that reality: conservative control laws, continuous monitoring, and automatic shutdowns when signals drift [40,41]. On the software side, "automation isn't automatic": deployment demands careful procedural assessment, instrument integration, and robust data handling before a single closed-loop run makes sense [42]. Where large language models intersect with wet chemistry, safety filters and human-in-the-loop checks are essential—hallucinated protocols or misread manuals are not theoretical risks [43].

6.2.3. Intellectual Property and Attribution in Human–AI Work

As autonomy increases, we need clarity about who (or what) counts as an inventor or author. Patent offices are actively probing whether an AI system can contribute to inventorship, particularly in joint human–AI scenarios [43]. Scientific publishing is already encountering fully AI-generated manuscripts—"AI Scientist-v2" demonstrated workshop-level automated discovery and paper drafting—which raises concrete questions about disclosure, authorship, and accountability [44].The immediate, pragmatic stance is transparency: disclose AI assistance, retain human responsibility for claims and data, and align with evolving journal and patent guidelines.

6.2.4. Rigor and Peer Review in an Automated Age

Machine-learning-guided chemistry is only as good as its data and its uncertainty estimates. Literature-mined reaction corpora are biased toward high-yield outcomes and may contain errors or inconsistent metadata; without negative results, models learn fuzzy decision boundaries [22]. Ultrahigh-throughput platforms and structured logging solve much of this by capturing successes and failures in machine-readable form, producing internally consistent, auditable datasets suited for modeling. Generative models add a different challenge: hallucinated text and opaque provenance. The remedy is twofold: algorithms trained on high-quality, peer-reviewed corpora and mandatory disclosure with human verification of any AI-generated content [39]. The stakes are nontrivial; even simple trials have shown that naive model outputs can cite incorrectly at alarming rates, underscoring the need for checks before publication. Peer review can adapt—by expecting uncertainty reporting, deployment-realistic validation (scaffold-wise, lot-wise), and explicit accounting of negative data—so that automated pipelines raise, rather than erode, standards [4,7,8].

Transition

The trajectory is clear. Quantum acceleration will sharpen the physics; digital twins will keep models synchronized with the world; immersive tools will make oversight natural; and carefully designed autonomy will scale what human chemists do best. None of this replaces judgment. It amplifies it—provided we invest in the skills, safety, attribution, and rigor that let predictive chemistry be trustworthy by default. In the conclusion, we bring the arc back to its simplest form: fewer surprises at the bench, more discoveries on purpose, and a generation of chemists fluent in both code and chemistry.

Conclusion

If there is one through-line to this paper, it is this: stop forcing a false choice. The calculation is not the enemy of the flask; the flask is not the enemy of the calculation. Chemistry only becomes

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

predictive when we let them argue in public, with machine learning acting as the translator that keeps the debate honest.

We began by naming the thing most of us have felt: the reality gap. In the places that matter most—bonds stretching and breaking, surfaces choosing a pathway, solvents shifting the free energy by just enough to flip an outcome, spin states changing the map as you walk it—our neat theories wobble. That's not a failure of science; it's a reminder that approximation has edges. What has changed is our ability to correct those edges systematically, not by throwing physics away, but by teaching it where it goes wrong and quantifying how wrong it is. Hybrid correctors do not replace first principles; they rescue them. Uncertainty stops being an apology and becomes a design variable. And when instruments speak in real time, models stop issuing proclamations and start making decisions.

This is what predictive chemistry looks like in practice. You keep the physics where it is strong. You learn the residual where it is stubbornly wrong. You carry uncertainty forward so every prediction comes with its own humility. You couple the whole thing to analytics that watch the reaction breathe. Then you let a Bayesian engine pick the next experiment, not to chase a single "best" number, but to advance the frontier between what you know and what you don't—across yield, selectivity, cost, and greenness. The loop does not end with a plot; it ends with a better reaction, a cleaner dataset, and a choice made with eyes open.

We also learned what "good" looks like when the work leaves the slide deck. Validation mirrors deployment, not convenience. Splits reflect how the chemistry will actually vary. Negative results are logged because they draw the line between "works" and "doesn't." Popular conditions are treated as starting points, not destiny. Autonomy is conditional by design, with clear thresholds where the system pauses and asks for judgment. And everything is encoded—not to feed a bureaucracy, but to make reproducibility the default and to leave a trail the next scientist can trust.

Why does this matter? Because time is the rarest reagent. Faster cycles and structured data mean fewer months lost to blind alleys. Because trust is the currency of adoption. Calibrated uncertainty and auditable decisions invite people to use the tool, not fear it. Because sustainability is no longer a side constraint. When you ask for greener routes up front, you get them. And because this changes our work for the better. Robots will not replace chemists; they will replace the parts of chemistry that deserve to be automated, so we can do the parts that make us scientists: choosing the right problems, framing testable ideas, and adjudicating the edge cases where craft still matters.

The path forward is refreshingly practical. Start where feedback is richest. Pair live analytics with optimization so each experiment teaches you something you didn't know. Use corrective models where physics is biased, and switch to simpler learners when the data are abundant. Validate as you intend to deploy. Treat failures as first-class data. Build your autonomy in layers and keep humans firmly on the loop. And teach this way of working—side by side with spectroscopy and synthesis—so the next generation is fluent in both code and glassware.

If we do that, the small heartbreak we began with does not disappear, but it does soften. The calculation and the flask won't always agree. They shouldn't. But they will disagree productively, more often, and for reasons we can understand. That is what it feels like when a field grows up: fewer surprises, more discoveries on purpose, and a lab culture that treats its tools not as rivals, but as partners. The bridge is here. Our job now is to cross it—together—and keep building as we go.

Conflicts of Interest: There are no conflicts to declare

References

1. Wang, Y.; Lin, Z.; Ouyang, R.; Jiang, B.; Zhang, I. Y.; Xu, X. A Renormalized Doubly Hybrid Method Enhanced with Machine Learning for a Unified Treatment of Static and Dynamic Correlations. ChemRxiv March 8, 2024. https://doi.org/10.26434/chemrxiv-2024-th0d9.
2. *Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities | Chemical Reviews*. https://pubs.acs.org/doi/10.1021/acs.chemrev.8b00361 (accessed 2025-07-24).

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

3. Schneebeli, S. T. Computers for Chemistry and Chemistry for Computers: From Computational Prediction of Reaction Selectivities to Novel Molecular Wires for Electrical Devices, Columbia University, 2011. https://doi.org/10.7916/D8V4125M.
4. Karande, P.; Gallagher, B.; Han, T. Y.-J. A Strategic Approach to Machine Learning for Material Science: How to Tackle Real-World Challenges and Avoid Pitfalls. *Chem. Mater.* **2022**, *34* (17), 7650–7665. https://doi.org/10.1021/acs.chemmater.2c01333.
5. Ishida, S.; Terayama, K.; Kojima, R.; Takasu, K.; Okuno, Y. AI-Driven Synthetic Route Design Incorporated with Retrosynthesis Knowledge. *J. Chem. Inf. Model.* **2022**, *62* (6), 1357–1367. https://doi.org/10.1021/acs.jcim.1c01074.
6. Luise, G.; Huang, C.-W.; Vogels, T.; Kooi, D.; Ehlert, S.; Lanius, S.; Giesbertz, K. J. H.; Karton, A.; Gunceler, D.; Stanley, M.; Bruinsma, W.; Huang, L.; Wei, X.; Torres, J. G.; Katbashev, A.; Zavaleta, R. C.; Máté, B.; Kaba, S.-O.; Sordillo, R.; Chen, Y.; Williams-Young, D. B.; Bishop, C.; Hermann, J.; Berg, R. van den; Gori-Giorgi, P. Accurate and Scalable Exchange-Correlation with Deep Learning. **2025**.
7. Proppe, J.; Kircher, J. Transforming Predictions into Testable Hypotheses: The Case of Polar Organic Reactivity. ChemRxiv February 26, 2021. https://doi.org/10.26434/chemrxiv.14102372.v1.
8. Proppe, J.; Kircher, J. Uncertainty Quantification of Reactivity Scales. *Chemphyschem* **2022**, *23* (8), e202200061. https://doi.org/10.1002/cphc.202200061.
9. Kuntz, D.; Wilson, A. K. Machine Learning, Artificial Intelligence, and Chemistry: How Smart Algorithms Are Reshaping Simulation and the Laboratory. *Pure and Applied Chemistry* **2022**, *94* (8), 1019–1054. https://doi.org/10.1515/pac-2022-0202.
10. Velasco, P. Q.; Hippalgaonkar, K.; Ramalingam, B. Emerging Trends in the Optimization of Organic Synthesis through High-Throughput Tools and Machine Learning. *Beilstein J Org Chem* **2025**, *21*, 10–38. https://doi.org/10.3762/bjoc.21.3.
11. Tobias, A. V.; Wahab, A. Autonomous 'Self-Driving' Laboratories: A Review of Technology and Policy Implications. *Royal Society Open Science* **2025**, *12* (7), 250646. https://doi.org/10.1098/rsos.250646.
12. Burger, B. A Mobile Robotic Researcher. dphil, University of Liverpool, 2020. https://livrepository.liverpool.ac.uk/3087073 (accessed 2025-07-22).
13. Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533* (7601). https://doi.org/10.1038/nature17439.
14. Li, Y.; [et al.]. *Transforming the Synthesis of Carbon Nanotubes with Machine Learning Models and Automation*. arXiv preprint arXiv:2404.01006. https://arxiv.org/abs/2404.01006.
15. Sans, V.; Porwol, L.; Dragone, V.; Cronin, L. A Self Optimizing Synthetic Organic Reactor System Using Real-Time in-Line NMR Spectroscopy. *Chem. Sci.* **2015**, *6* (2), 1258–1264. https://doi.org/10.1039/C4SC03075C.
16. Malashin, I.; Tynchenko, V.; Gantimurov, A.; Nelyub, V.; Borodulin, A. Physics-Informed Neural Networks in Polymers: A Review. *Polymers* **2025**, *17* (8), 1108. https://doi.org/10.3390/polym17081108.
17. Yahagi, Y.; Obuchi, K.; Kosaka, F.; Matsui, K. Transfer Learning from First-Principles Calculations to Experiments with Chemistry-Informed Domain Transformation. arXiv April 7, 2025. https://doi.org/10.48550/arXiv.2504.02848.
18. Teixeira, R. I.; Andresini, M.; Luisi, R.; Benyahia, B. Computer-Aided Retrosynthesis for Greener and Optimal Total Synthesis of a Helicase-Primase Inhibitor Active Pharmaceutical Ingredient. *JACS Au* **2024**, *4* (11), 4263–4272. https://doi.org/10.1021/jacsau.4c00624.
19. Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D. Machine Learning Meets Mechanistic Modelling for Accurate Prediction of Experimental Activation Energies. *Chem. Sci.* **2021**, *12* (3), 1163–1175. https://doi.org/10.1039/D0SC04896H.
20. Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4* (11), 1465–1476. https://doi.org/10.1021/acscentsci.8b00357.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

21. Shields, J. D.; Howells, R.; Lamont, G.; Leilei, Y.; Madin, A.; Reimann, C. E.; Rezaei, H.; Reuillon, T.; Smith, B.; Thomson, C.; Zheng, Y.; Ziegler, R. E. AiZynth Impact on Medicinal Chemistry Practice at AstraZeneca. *RSC Med. Chem.* **2024**, *15* (4), 1085–1095. https://doi.org/10.1039/D3MD00651D.
22. Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54* (10), 2337–2346. https://doi.org/10.1021/acs.accounts.1c00119.
23. Sawicki, J. W.; Bogdan, A. R.; Searle, P. A.; Talaty, N.; Djuric, S. W. Rapid Analytical Characterization of High-Throughput Chemistry Screens Utilizing Desorption Electrospray Ionization Mass Spectrometry. *React. Chem. Eng.* **2019**, *4* (9), 1589–1594. https://doi.org/10.1039/C9RE00054B.
24. Logsdon, D. L.; Li, Y.; Paschoal Sobreira, T. J.; Ferreira, C. R.; Thompson, D. H.; Cooks, R. G. High-Throughput Screening of Reductive Amination Reactions Using Desorption Electrospray Ionization Mass Spectrometry. *Org. Process Res. Dev.* **2020**, *24* (9), 1647–1657. https://doi.org/10.1021/acs.oprd.0c00230.
25. Wleklinski, M.; Loren, B. P.; Ferreira, C. R.; Jaman, Z.; Avramova, L.; Sobreira, T. J. P.; Thompson, D. H.; Cooks, R. G. High Throughput Reaction Screening Using Desorption Electrospray Ionization Mass Spectrometry. *Chem. Sci.* **2018**, *9* (6), 1647–1653. https://doi.org/10.1039/C7SC04606E.
26. Sinclair, I.; Bachman, M.; Addison, D.; Rohman, M.; Murray, D. C.; Davies, G.; Mouchet, E.; Tonge, M. E.; Stearns, R. G.; Ghislain, L.; Datwani, S. S.; Majlof, L.; Hall, E.; Jones, G. R.; Hoyes, E.; Olechno, J.; Ellson, R. N.; Barran, P. E.; Pringle, S. D.; Morris, M. R.; Wingfield, J. Acoustic Mist Ionization Platform for Direct and Contactless Ultrahigh-Throughput Mass Spectrometry Analysis of Liquid Samples. *Anal. Chem.* **2019**, *91* (6), 3790–3794. https://doi.org/10.1021/acs.analchem.9b00142.
27. Avila, C.; Cassani, C.; Kogej, T.; Mazuela, J.; Sarda, S.; Clayton, A. D.; Kossenjans, M.; Green, C. P.; Bourne, R. A. Automated Stopped-Flow Library Synthesis for Rapid Optimisation and Machine Learning Directed Experimentation. *Chem. Sci.* **2022**, *13* (41), 12087–12099. https://doi.org/10.1039/D2SC03016K.
28. *Where is the return on investment (ROI) from lab automation? - Synthace*. https://www.synthace.com/blog/where-is-the-return-on-investment-roi-from-lab-automation (accessed 2025-10-05).
29. Reizman, B. J.; Jensen, K. F. Feedback in Flow for Accelerated Reaction Development. *Acc. Chem. Res.* **2016**, *49* (9), 1786–1796. https://doi.org/10.1021/acs.accounts.6b00261.
30. *Economic potential of generative AI | McKinsey*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier (accessed 2025-10-05).
31. Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **2017**, *50* (12), 2976–2985. https://doi.org/10.1021/acs.accounts.7b00428.
32. Evangelista, F. A.; Batista, V. S. Editorial: Quantum Computing for Chemistry. *J. Chem. Theory Comput.* **2023**, *19* (21), 7435–7436. https://doi.org/10.1021/acs.jctc.3c01043.
33. Peterson, L.; Gosea, I. V.; Benner, P.; Sundmacher, K. Digital Twins in Process Engineering: An Overview on Computational and Numerical Methods. *Computers & Chemical Engineering* **2025**, *193*, 108917. https://doi.org/10.1016/j.compchemeng.2024.108917.
34. Eyke, N. S.; Green, W. H.; Jensen, K. F. Iterative Experimental Design Based on Active Machine Learning Reduces the Experimental Burden Associated with Reaction Screening. *React. Chem. Eng.* **2020**, *5* (10), 1963–1972. https://doi.org/10.1039/D0RE00232A.
35. Fombona-Pascual, A.; Fombona, J.; Vicente, R. Augmented Reality, a Review of a Way to Represent and Manipulate 3D Chemical Structures. *J Chem Inf Model* **2022**, *62* (8), 1863–1872. https://doi.org/10.1021/acs.jcim.1c01255.
36. Seifrid, M.; Pollice, R.; Aguilar-Granda, A.; Morgan Chan, Z.; Hotta, K.; Ser, C. T.; Vestfrid, J.; Wu, T. C.; Aspuru-Guzik, A. Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab. *Acc Chem Res* **2022**, *55* (17), 2454–2466. https://doi.org/10.1021/acs.accounts.2c00220.
37. *dspace.mit.edu/handle/1721.1/144518*. https://dspace.mit.edu/handle/1721.1/144518 (accessed 2025-10-05).
38. Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023**, *123* (6), 3089–3126. https://doi.org/10.1021/acs.chemrev.2c00798.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 October 2025 doi:10.20944/preprints202510.1185.v1

39. Blanco-González, A.; Cabezón, A.; Seco-González, A.; Conde-Torres, D.; Antelo-Riveiro, P.; Piñeiro, Á.; Garcia-Fandino, R. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals (Basel)* **2023**, *16* (6), 891. https://doi.org/10.3390/ph16060891.
40. Yang, Q.; Canturk, B.; Gray, K.; McCusker, E.; Sheng, M.; Li, F. Evaluation of Potential Safety Hazards Associated with the Suzuki–Miyaura Cross-Coupling of Aryl Bromides with Vinylboron Species. *Org. Process Res. Dev.* **2018**, *22* (3), 351–359. https://doi.org/10.1021/acs.oprd.8b00001.
41. Yang, Q.; Babij, N. R.; Good, S. Potential Safety Hazards Associated with Pd-Catalyzed Cross-Coupling Reactions. *Org. Process Res. Dev.* **2019**, *23* (12), 2608–2626. https://doi.org/10.1021/acs.oprd.9b00377.
42. Christensen, M.; Yunker, L. P. E.; Shiri, P.; Zepel, T.; Prieto, P. L.; Grunert, S.; Bork, F.; Hein, J. E. Automation Isn't Automatic. *Chem. Sci.* **2021**, *12* (47), 15473–15490. https://doi.org/10.1039/D1SC04588A.
43. Kim, D. 'AI-Generated Inventions': Time to Get the Record Straight? *GRUR Int* **2020**, *69* (5), 443–456. https://doi.org/10.1093/grurint/ikaa061.
44. Yamada, Y.; Lange, R. T.; Lu, C.; Hu, S.; Lu, C.; Foerster, J.; Clune, J.; Ha, D. The AI Scientist-v2: Workshop-Level Automated Scientific Discovery via Agentic Tree Search. arXiv April 10, 2025. https://doi.org/10.48550/arXiv.2504.08066.

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

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.