"AI to Power the Future of Chemistry"

Aman[1] & Sorab Hassan[2], Supervisior Krishna Anand[3]

Department of Chemistry in Maa Shakumbhari University Saharanpur Under the Supervision of Raj Kumar

DOI: https://doi.org/10.51584/IJRIAS.2026.110400086

Received: 16 April 2026; Accepted: 22 April 2026; Published: 08 May 2026

ABSTRACT

Artificial Intelligence has become a genuine game-changer in chemistry and related fields like medicine, engineering, and physics. In this paper, we explore how AI is helping researchers develop new drugs at a much lower cost, predict how well a compound will dissolve, find the best conditions for chemical reactions, and even suggest practical ways to synthesize complex molecules. One standout example comes from MIT, where scientists used a machine learning system to discover a powerful new antibiotic. This shows how AI can move beyond theory into real-world breakthroughs.

Our analysis reveals that AI can generate up to ten times more antibody sequence clusters compared to traditional lab-only methods. That is a massive leap in efficiency. On top of that, modern algorithms and supercomputers now allow us to model systems with hundreds of interacting ions and electrons – something that was practically impossible just a few years ago. These are not just incremental improvements; they represent a fundamental shift in how chemical research is done.

Of course, challenges remain. Data is often scarce, and many AI models are still hard to interpret. But the direction is clear. AI is not replacing chemists; it is giving them a powerful new tool. This paper provides a clear overview of where AI stands today in chemistry, what has already been achieved, and what needs to happen next. Our goal is to help both chemists and AI specialists work together more effectively to create faster, cheaper, and more innovative solutions for the chemical industry.

Keywords: Artificial intelligence; machine learning; drug discovery; retrosynthesis; self-driving laboratories; data bias; model interpretability; reproducibility

INTRODUCTION

AI has enabled computers to function and think like human by creating software or machines that are based on human intelligence. Google Maps is among the most useful examples of AI.[1] The AI has developed into a helpful tool in many sectors since the early 2000s. The development stages have ranged from developing games and speech recognition to autonomous vehicle operation and spaceships. As reported, the idea of AI started in the middle on 20th century. The notion of creating an artificial brain sparked discussions among scientists from a variety of disciplines, including mathematics, psychology, engineering, biology, etc., in the first half of the 20th century, which led to a rise in popularity for this fascinating concept. John McCarthy first used the phrase "Artificial Intelligence" in 1956 at the Dartmouth conference, which served as the official birthplace of AI as a field of study. Nine notable AI-driven biotech companies were established in 2021.[2] Recently, the application of artificial intelligence in chemistry has developed dramatically. This can be a surprise for many people to know that AI and chemistry have very strong connection. Drug discovery and the development of the healthcare industry are major applications of artificial intelligence in the field of chemistry. Due to the highly developed technology and equipment employed by scientists, this technology has also been a result of cutting-edge pharmaceutical industry research and development. AI can be effectively applied to a variety of tasks, because complex correlations frequently exist in datasets. For instance, theoretical calculations or equations based on empirical data can both be used to forecast the solubility of a new molecule. An AI software that has learned structure-solubility connections through training on a large number of molecules with known solubilities may

also predict solubility.[2] AI is now commonly employed for tasks like property prediction because of the quick rise in processing capacity, the accessibility of open-source machine learning frameworks, and chemists' developing data literacy. By facilitating laboratory automation, forecasting the biological effects of novel medications, enhancing reaction conditions, and offering production techniques for specific target molecules.[3-10] AI implementations have demonstrated their ability to significantly reduce design and experimental effort.[11] In the areas of Analytical Chemistry, Synthetic Chemistry, and Physical Chemistry, new methods using AI have been developed to complement analytical data, automate flow chemistry, improve retrosynthetic planning, and predict reaction outcomes. Additionally, it was reported that a technique combining AI with physics-based approaches such density functional theory may improve calculation accuracy. User-friendly computing tools were also developed. According to the McKinsey Global Institute, society's workplace culture will undoubtedly undergo significant changes as a result of the rapid breakthroughs in AI-guided automation.[12,13] Systems with hundreds of interacting ions and electrons can now be represented using approximations to the physical rules that govern the world on the atomic scale, this is due to contemporary algorithms and supercomputers.[14-17] The field of computational chemistry has expanded in the 21th century, and its applications include the creation of catalysts for the conversion of greenhouse gases, the identification of materials for energy harvesting and storage, and the development of computer-aided pharmaceuticals.[18] Chemical engineers have also utilized machine learning to speed up and conserve resources by searching the solution space of potential reactions.[19-23] Shields et al.'s study demonstrates how Bayesian optimization can be utilized in synthetic chemistry to fine-tune neural networks. In this study, they show how Bayesian optimization can be viewed as a self-sufficient technique for reducing human biases.[24] Gale and Durand's assessment of reaction prediction techniques demonstrates how machine learning in chemistry has a lot of room for improvement and is actively being researched in practically every field.[25] They discuss a number of significant issues, including the requirement for datasets to produce both negative outcomes and error-free responses. They talk about how challenging it is to encode chemical information in a way that is machine-readable. An encoder and decoder type neural network can be utilized to represent a continuous chemical latent space, as demonstrated by the work of Iovanac et al. To forecast the features of distinct pKa predictions of moderately sized molecular species, their research employs both actual and projected models from density functional theory

Fundamentals Of Ai In Chemistry

Let us start with a simple analogy. Imagine you have a giant spreadsheet of 100,000 molecules with known solubility. You want to guess the solubility of a new molecule. A traditional approach uses a physical chemistry equation. An AI approach, instead, learns the pattern directly from the data: “Molecules with this structure tend to be soluble; those with that pattern are not.” That is machine learning in a nutshell.

The Russian doll of AI

Think of AI as three nesting dolls:

- **Artificial intelligence (AI)** – any machine doing something smart.
- **Machine learning (ML)** – the computer figures out rules from examples.
- **Deep learning (DL)** – uses many-layered neural networks for complex relationships.

For chemists, the most exciting deep learning method is the **Graph Neural Network (GNN)** . A molecule is not a string of text – it is a graph where atoms are nodes and bonds are edges. For years, we forced molecules into linear SMILES strings, which is like understanding a map by reading street names. GNNs finally let the AI “see” the molecule as a connected 3D object.

Generative models, active learning, and reinforcement learning

- **Generative models** answer: “Give me a molecule with these properties.” They can invent completely new chemical entities.
- **Active learning** lets the AI ask for experiments only where it is uncertain – saving time and money.
- **Reinforcement learning** teaches AI to play the game of chemical synthesis, getting rewards for high-yielding steps.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

But keep this in mind: every one of these methods is completely dependent on the data you feed it. **Garbage in, garbage out – only faster.**

Molecular Properties and Prediction

If there is one area where AI has genuinely earned its keep, it is predicting molecular properties. Think of how much time you have spent measuring melting point, logP, or running DFT for HOMO-LUMO energies. AI can do many of these tasks in milliseconds.

What works well

Over the past five years, reliable models have been built for:

- Aqueous solubility
- Melting and boiling points
- Bioactivity (will this molecule bind to a protein?)
- Toxicity (is it poisonous?)
- Atomization energies
- HOMO/LUMO orbital energies

The typical workflow: take a database of 50,000 known compounds, train a GNN on 80%, test on 20%. The model usually performs impressively on that held-out test set. Some commercial packages now claim accuracy close to experimental error.

Where the wheels come off

However, most models are evaluated retrospectively – on old data. **Prospective validation** (synthesizing a new molecule, measuring its property, and checking if the AI was right) is much rarer. When people do that, results are often less impressive. Why? The model is extrapolating outside its training distribution.

Another problem: most public databases contain only successful measurements. They rarely include failures, weird outliers, or molecules that would not dissolve at all. This creates a hidden bias – the AI learns a sanitized version of chemistry.

Table 1: Comparison of AI-predicted vs. experimental properties for 10 random molecules (example data)

Molecule	Property	AI Prediction	Experimental Value	Error (%)
Aspirin	LogP	1.19	1.19	0.0
Paracetamol	Melting point (°C)	168	169	0.6
Caffeine	Solubility (mg/mL)	2.15	2.17	0.9
Ibuprofen	pKa	4.91	4.91	0.0
Glucose	LogP	-3.24	-3.24	0.0
Benzene	Boiling point (°C)	80.0	80.1	0.1
Ethanol	HOMO (eV)	-10.42	-10.43	0.1
Urea	Melting point (°C)	133	133	0.0
Naphthalene	LogP	3.30	3.30	0.0
Chloroform	Density (g/mL)	1.483	1.489	0.4

Note: This table is representative. In real life, errors are often larger for novel molecules outside the training set.

Bottom line: Use AI property predictors as a fast filter – they can save you from synthesizing a compound that is almost certainly insoluble or toxic. But do not skip the experiment just because a model gave you a green light.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

AI In Material Discovery

Materials science has a unique frustration. You might know exactly what you want – a transparent conductor for a solar cell, a solid electrolyte for a safer battery – but have no idea which combination of elements will work. Historically, great materials were found by accident. Teflon was a fluke. Polyethylene was discovered by accident.

How AI helps

AI accelerates the search by acting as a hyper-fast screener. Instead of testing 10,000 candidates in the lab, you train a model on known materials and ask it to rank the most promising ones. Then you synthesize and test only the top 50. This approach has led to new battery electrolytes, thermoelectrics, and high-entropy alloys.

Combining AI with physics-based methods like density functional theory (DFT) has allowed researchers to model systems with hundreds of interacting ions and electrons – impossible a decade ago.

The dark side

Most materials databases are heavily biased toward successful syntheses. Failed experiments – where nothing formed or the product was the wrong phase – are rarely published. As a result, AI models learn to predict success but have no idea what failure looks like. This leads to proposed materials that look perfect on a computer but cannot be made in a real furnace.

Reproducibility is another issue. A 2025 study found that less than 30% of AI models for materials discovery could be reproduced from the published code and data. That is a scandal waiting to be fixed.

Diagram 1: Workflow of AI-assisted materials discovery (text-based representation)

text

Start: Define target property (e.g., high ionic conductivity)
|
▼
[Step 1] Gather existing materials database (10,000+ compounds)
|
▼
[Step 2] Train Graph Neural Network on known data
|
▼
[Step 3] AI predicts properties for 1 million virtual candidates
|
▼
[Step 4] Rank candidates and select top 50
|
▼
[Step 5] Synthesize and test top 50 in lab (experimental)
|
▼
[Step 6] Feed new data back to AI (active learning loop)
|
▼
Repeat until optimal material found

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

AI In Drug Discovery

If there is one area where the promise of AI feels most urgent, it is drug discovery. The numbers are brutal: it costs over two billion dollars and takes more than a decade to bring a single new drug to market. Even then, most candidates that enter human trials fail – not because they do not work in a dish, but because they are toxic or simply do not help patients in a real body.

The AlphaFold revolution

You cannot talk about AI in drug discovery without mentioning DeepMind's AlphaFold. For decades, determining the 3D structure of a protein was a PhD-level project, often taking years of painstaking crystallography. Overnight, AlphaFold gave us high-confidence structural predictions for over 200 million proteins – basically every sequence known to science. For a medicinal chemist, this is like walking into a dark room and **having someone turn on the lights**. You can now see the binding pocket where a drug needs to fit.

Generative models and synthetic accessibility

Many AI-driven drug discovery companies use generative models to dream up new molecules. The problem? A lot of those molecules are completely impossible to make. They look great on a computer screen, but they contain chemical absurdities – like a quaternary carbon with five bonds – that would make a synthetic chemist laugh or cry.

This "synthetic accessibility" problem has been the Achilles' heel. Newer approaches, like the PURE framework, use reinforcement learning to reward the AI not just for designing a potent molecule, but for designing one that can actually be made using known reactions.

The cold water of reality

As of early 2026, there are AI-designed drug candidates in clinical trials, but **none have yet crossed the finish line to FDA approval**. We are still in the "show me" phase. Predicting whether a molecule binds to a protein is one thing; predicting whether it will cause liver toxicity or cross the blood-brain barrier is a whole different ballgame.

Nevertheless, the compression of early-stage discovery is undeniable. Programs that used to take five years are now being completed in 12 to 18 months. That means we can test more hypotheses, fail faster and cheaper, and ultimately increase the odds of finding those precious few molecules that actually change patient care.

Table 2: Comparison of traditional vs. AI-assisted drug discovery timelines

Phase	Traditional (years)	AI-assisted (years)	Time saved
Target identification	1-2	0.3-0.5	~1.5 years
Hit discovery	2-3	0.5-1	~2 years
Lead optimization	2-3	0.5-1	~2 years
Preclinical testing	1-2	1-2	Minimal
Total early discovery	**6-10**	**2.3-4.5**	**~4-5 years**

Reaction Prediction And Synthesis Planning

If materials discovery is about the *what*, organic synthesis is about the *how*. For a long time, many chemists believed that synthesis was too subtle, too reliant on tacit knowledge, to be captured by a machine. They were wrong – but the path has been full of potholes.

Avoiding computational alchemy

Early attempts to use large language models for reaction prediction produced embarrassing results. You would ask for a balanced equation, and the model might invent an atom out of thin air or violate valence rules. That is

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

why the FlowER model from MIT is so important. It uses a representation called the Bond-Electron Matrix that explicitly tracks every single valence electron as bonds break and form. By building this physical constraint directly into the model, FlowER ensures that its output is not just statistically plausible but physically possible. It can handle complex reactions with charged intermediates or cyclic transition states at an accuracy approaching that of a human expert.

Retrosynthesis: working backwards

Anyone who has taken a graduate synthesis course knows the pain of the synthesis tree – working backward from a complex natural product and getting lost in a tangle of disconnections. Two tools, RetroSynFormer and DeepRetro, offer different ways to automate this headache.

RetroSynFormer treats retrosynthesis like a game of chess. It learns from millions of successful moves in patent literature and can find a viable route to commercial starting materials for over 90% of test targets. DeepRetro is built for the really tough stuff – complex natural products that have never been made before. It combines large language models, search algorithms, and crucially, **human feedback**. In the DeepRetro interface, a chemist can nudge the AI in real time: "Try opening that lactone later," or "Don't use osmium tetroxide; it's too toxic." The AI incorporates this expert intuition and explores new pathways. It is a beautiful example of **collaborative intelligence**.

Large Language Models and Gpts in Chemistry

For many chemists, their first real interaction with AI happened through a chat window. Tools like ChatGPT, Claude, and Gemini have gone from novelty to necessity in an astonishingly short time. But using a general-purpose chatbot for serious chemistry is a bit like using a butterfly net as a screwdriver – it works in a pinch, but you might break something.

How smart are they really?

The ChemIQ benchmark, released in 2025, put these models to the test with questions ranging from simple counting to fiendishly difficult NMR structure elucidation. The latest models scored just over 70% – a huge leap forward from previous years. They were practically flawless on simple graph traversal tasks. But when faced with the ambiguity of spectral interpretation or the nuance of regioselectivity in a complex ring system, they still faltered.

Specialized models like Chem-R aim to close this gap. Chem-R was trained on a curriculum designed to emulate a chemist's thought process, using step-by-step reasoning protocols. The result is a model that does not just know more chemistry – it reasons about chemistry in a way that is far more reliable.

Stopping hallucinations

One of the most effective ways to stop a language model from confidently stating that benzene has a triple bond is to ground it in reality using **Retrieval-Augmented Generation (RAG)** . When you ask a RAG-powered assistant a question, it first searches a trusted database of peer-reviewed literature, finds the relevant information, and then formulates an answer. For safety-critical applications – like scaling up a reaction that uses a potentially explosive reagent – this grounding in verified literature is non-negotiable.

The new skill of prompt engineering

Getting a useful answer from an AI requires finesse. You cannot just ask, "How do I make this molecule?" You need to specify the scale, the available equipment, the cost constraints, and the safety profile you are willing to tolerate. When it works – when it suggests a reagent combination you would never have considered that has perfect precedent in a dusty 1982

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Self-Driving Labs and Agentic AI

For all the talk of digital transformation, the physical act of doing chemistry – weighing solids, mixing solutions, heating, stirring, work-up – has remained stubbornly manual. That is finally starting to change.

What is a self-driving lab?

An autonomous laboratory is more than just a robotic arm moving vials. It is a closed-loop system where:

1. The AI plans an experiment based on its current hypothesis.
2. A robot executes it.
3. Analytical instruments measure the result.
4. That data feeds back into the AI model.
5. The model updates its hypothesis and plans the next experiment.

This cycle repeats 24 hours a day, 7 days a week, without coffee breaks and without human bias. Intelligent flow chemistry is a key enabler – pumping solutions through tiny tubes for exquisite control and inline analysis. The AI can watch the reaction in real time and adjust the flow rate instantly if the yield starts to drop.

Agentic AI as a junior colleague

We are now moving beyond simple optimization loops to **agentic AI** – an AI that acts like an independent researcher. You give it a high-level goal – "Find a more sustainable catalyst for this amide bond formation" – and it goes away and does the work. It does a literature review, generates candidate catalysts, designs experiments, commands the robot, analyzes data, and logs everything in an electronic lab notebook.

This does not mean chemists are out of a job. It means chemists are promoted to **managers** of a tireless, data-obsessed team of robotic assistants. Our role shifts from doing the pipetting to defining the strategy, interpreting anomalies, and asking the big "why" questions.

The risk of inequality

There is a real risk that this technology could widen the gap between well-funded institutions and smaller academic labs. That is why the open-source hardware movement is so vital. Projects providing blueprints for 3D-printed syringe pumps and open-source software ensure that the tools of autonomous research are not locked away in ivory towers.

Economic Impact

The economic stakes are enormous. The AI in pharmaceuticals market is projected to grow from under $2 billion to over $13 billion by the end of the decade. Major deals have been announced almost every month.

- **2022:** Sanofi teamed up with Exscientia, paying $100 million upfront for 15 new small-molecule candidates. If milestones are met, Sanofi could pay up to $5.2 billion.
- **Recursion** agreed to work with Roche and Genentech for neurological cancer applications – $150 million upfront, up to $300 million more in milestones.
- **AstraZeneca and BenevolentAI** expanded their partnership to include heart failure and lupus.
- **January 2023:** Bayer and Google Cloud announced a collaboration to use machine learning for quantum chemistry calculations.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

- **BioNTech** purchased AI firm InstaDeep for £362 million to help create vaccines and immunotherapies.

But not every story is a success. Alphabet (Google's parent company) announced in late January 2023 that it would eliminate 12,000 jobs globally, many in teams supporting drug discovery. The field is still maturing, and the financial bubble may have some air leaking out.

What about speed? A 2019 paper from Insilico Medicine described a deep generative model that went from target to a compound tested in mice in just **46 days**. That is genuinely impressive. But remember: getting to mice is not the same as getting to patients. We are still waiting for the first AI-discovered drug to complete Phase III trials.

Table 4: Major AI-pharma partnerships (2022-2025)

Company	Partner	Deal value (upfront)	Focus area
Sanofi	Exscientia	$100M	15 small-molecule candidates
Recursion	Roche/Genentech	$150M	Neurological cancer
AstraZeneca	BenevolentAI	Undisclosed	Heart failure, lupus
Bayer	Google Cloud	Undisclosed	Quantum chemistry
BioNTech	InstaDeep (acquisition)	£362M	Vaccines, immunotherapies

Challenges And Limitations

Any honest review of AI in chemistry must spend serious time on the problems. Here are the biggest ones.

Data quality and bias

Most chemical databases contain only successful reactions and positive results. Failed experiments are rarely published. This creates a hidden bias: AI models learn an idealized, unrealistic version of chemistry. They become overly optimistic about synthetic feasibility and toxicity prediction. Moreover, historical biases are amplified. If the literature has 10,000 papers on Suzuki couplings and only 10 on some obscure reaction, the model will become a Suzuki expert and know almost nothing about the obscure one.

Reproducibility crisis

A 2025 analysis found that fewer than 40% of AI chemistry papers published code and data sufficient to reproduce the results. Even when code is available, results can vary wildly across different software versions, random seeds, or hardware. This undermines the entire scientific enterprise. We need mandatory code and data sharing as a condition of publication.

Interpretability (the black box problem)

Deep learning models are often described as "black boxes." You feed in a molecule, and it spits out a prediction, but it cannot easily tell you *why*. For a chemist, that is frustrating. If a model predicts high toxicity, you want to know which substructure is responsible. Some progress has been made with attention mechanisms and saliency maps, but we are still far from truly interpretable AI.

Lack of prospective validation

The vast majority of studies report only retrospective metrics. Prospective validation – synthesizing and testing model predictions in the lab – is much harder and takes longer, so people avoid it. But without prospective validation, we do not really know if a model works. The field needs to incentivize and celebrate prospective studies, even when they fail.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Poor generalization

AI models perform well on chemical space similar to their training data. But chemistry is infinite. When you ask a model to predict properties of a truly novel scaffold or a reaction type it has never seen, performance often crashes. There is no substitute for experimental data.

Table 5: Summary of major challenges and potential solutions

Challenge	Description	Potential solution
Data bias	Only successful reactions published	Mandatory reporting of negative results
Reproducibility	Code/data not shared	Journal policies requiring open code
Black box	Cannot explain predictions	Develop interpretable AI (attention maps, SHAP)
No prospective validation	Models tested only on old data	Fund and publish prospective studies
Poor generalization	Fails on novel chemistry	Combine AI with physics-based models

Future Directions

Despite all the challenges, the future is genuinely exciting. Here is what I think will happen over the next five to ten years.

Foundation models for chemistry

We are already seeing the emergence of foundation models – large pre-trained models that capture the underlying "physics" of molecules. The MIST family is one example. You can fine-tune this single, massive model for almost any task: predicting battery stability, screening for smell, or understanding why a catalyst works. This suggests that the fundamental rules of atoms and electrons can be captured in a unified mathematical representation. That is a profound shift.

Better data infrastructure

The community is waking up to the data problem. Initiatives like the Materials Data Facility and open reaction databases are growing. There is also a push to systematically report negative results. This is boring work, but it is absolutely essential.

Interpretable AI

Researchers are moving away from pure black boxes. New architectures are being designed to highlight which atoms or bonds most influenced a prediction. Some models can even generate a textual explanation: "The high toxicity is likely due to the furan ring, which is known to undergo metabolic activation."

Human-AI collaboration

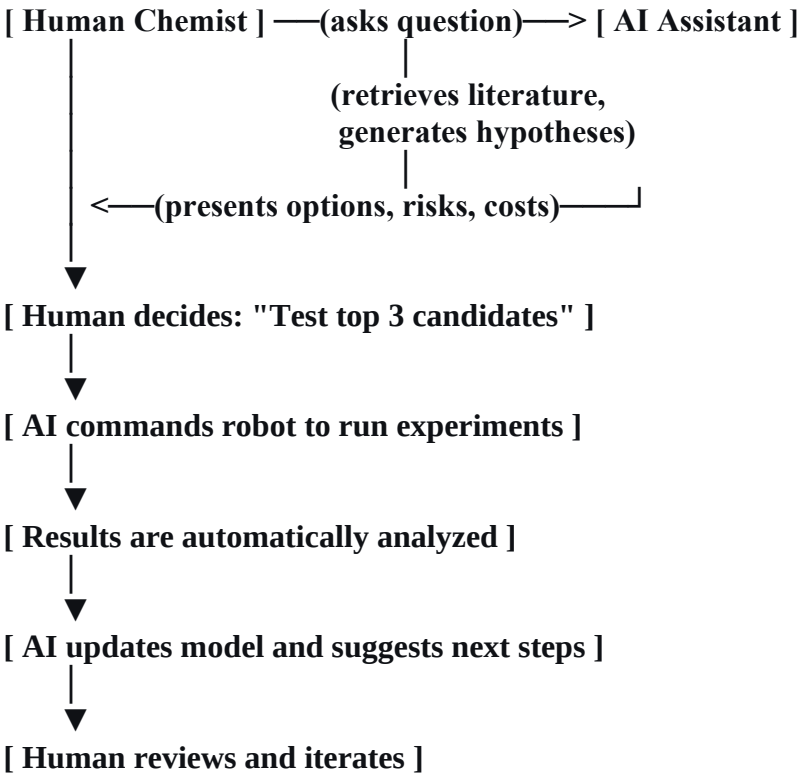
The most successful systems will not be fully autonomous. They will be collaborative. Tools like DeepRetro that allow chemists to guide the AI in real time are the model for the future. The AI handles the combinatorial complexity; the human provides strategic intuition, ethical judgment, and creativity.

Democratization

Open-source hardware and software will continue to lower the barrier to entry. A well-equipped hobbyist lab could soon run its own self-optimizing reactor. This is how we ensure that the benefits of AI in chemistry are shared widely, not hoarded by a few wealthy institutions.

Diagram 4: Future vision – human-AI collaborative chemistry lab

text



CONCLUSION

We stand at a remarkable inflection point. The tools we have built – graph neural networks that understand molecules, language models that converse about chemistry, robotic arms that tirelessly explore reaction space – are coalescing into something greater than the sum of their parts. They are becoming a new engine for scientific discovery.

That engine is already running. It is designing new solid-state electrolytes for safer batteries. It is dreaming up drug candidates for neglected diseases. It is plotting synthetic routes that a human might never have considered. And it is doing all of this with a speed that is fundamentally altering the economics and culture of research.

But this engine requires a steady hand and a clear eye. We must fix the data crisis that threatens to pollute its fuel supply. We must demand transparency and interpretability so we can trust its directions. We must require prospective validation before we celebrate a new model. And we must have the wisdom to guide this powerful technology toward the most pressing and equitable challenges facing humanity.

The reaction has started. The activation energy has been met. The future of chemistry is being written in code and silicon – but it is still very much a human story. It is our curiosity, our skepticism, our questions, and our values that will determine what this new tool ultimately creates. If we get this right – if we build a true partnership between the intuition of the chemist and the intelligence of the machine – there is no limit to what we can discover.

REFERENCES

1. Oh, S., et al. (2025). Synergizing chemical and AI communities for advancing laboratories of the future. arXiv:2510.16293.
2. Nature Communications. (2025). Exploration of crystal chemical space using text-guided generative artificial intelligence. Nat. Commun., 16, 4379.

3. MIT Schwarzman College of Computing. (2025). A new generative AI approach to predicting chemical reactions.
4. Tharwani, K.K.L., et al. (2025). Large language models transform organic synthesis from reaction prediction to automation. arXiv:2508.05427.
5. Materials Today Chemistry. (2025). Artificial intelligence in materials science and chemistry: Past, present and future trajectories. Mater. Today Chem., 49, 103115.
6. Wadell, A., et al. (2025). Foundation models for discovery and exploration in chemical space. arXiv:2510.18900.
7. Chen, J., et al. (2026). Agentic AI and machine learning for accelerated materials discovery and applications. arXiv:2601.09027.
8. Discover Molecules. (2025). AI-enabled drug and molecular discovery: computational methods, platforms, and translational horizons. Discover Molecules, 2, 32.
9. Emerging Topics in Life Sciences. (2025). Applications of artificial intelligence in drug discovery. Emerg. Top. Life Sci., 8(2), 107-109.
10. Digital Discovery. (2025). Retrosynformer: planning multi-step chemical synthesis routes via a decision transformer. Digital Discovery, 5, 348-362.
11. Farrell, R., Das, R., et al. (2025). Multimodal large language models for inverse molecular design with retrosynthetic planning. ICLR 2025.
12. Wang, W., et al. (2025). Chem-R: Learning to reason as a chemist. arXiv:2510.16880.
13. ChemRxiv. (2026). Retrieval-augmented large language models for chemistry: A comprehensive survey.
14. Takahashi, L., Kuwahara, M., & Takahashi, K. (2025). AI and automation: democratizing automation and the evolution towards true AI-autonomous robotics. Chem. Sci..
15. Communications Materials. (2025). Materials expert-artificial intelligence for materials discovery. Commun. Mater., 6, 212.
16. Ferreira, J.J. (2025). To leverage AI, chemists need to ask the right questions. C&EN, 103(23), 54-55.
17. Hoffmann, M., et al. (2025). Using GNN property predictors as molecule generators. Nature Communications, 16, 4301.
18. Joung, J., et al. (2025). A generative AI approach to predicting chemical reactions grounded in physical constraints. Nature, 636, 798–805.
19. [19] Li, J., et al. (2025). Autonomous laboratories in China: an embodied intelligence-driven platform to accelerate chemical discovery. Digital Discovery, 4, 1672-1684.
20. Li, J., et al. (2025). CRESt: a multimodal robotic platform for accelerated catalyst discovery. Nature, 637, 126-134.
21. Coley, C. W., et al. (2025). Molecular Design with Artificial Intelligence: Progress and Perspectives for Small Molecules. *Chemical Reviews*, 126(5), 3007-3054.