

Perspectives

Beyond Prediction: AI as a Creative Partner in Organic Synthesis

Tuan B. Vinh¹, Tao Li², Carl Yang², Monika Raj¹

Affiliation addresses are listed at the end of the article.

ABSTRACT

Artificial intelligence (AI) has rapidly progressed from being a niche curiosity to becoming a real force in the chemical sciences. In organic synthesis, however, AI's role is still mostly viewed in two ways: helping automate retrosynthetic planning and speeding up reaction optimization. These advances—like the retrosynthesis engine developed by Segler et al. (2018)¹ and the optimization workflows from Doyle's group (Shields et al., 2021)²—have been game-changing in specific cases. But if we see AI only in these roles, we risk overlooking its bigger potential.

Keywords Machine learning, Artificial intelligence, Organic synthesis, Reactivity and mechanisms, Cross-domain strategies, Chemical space expansion

received August 13, 2025 | accepted after revision September 30, 2025 | article published online 2025

Bibliography Synlett DOI 10.1055/a-2720-1562 Art ID ST-2025-08-0341-PE

© 2025. Thieme. All rights reserved. Georg Thieme Verlag KG, Oswald-Hesse-Straße 50, 70469 Stuttgart, Germany.

Correspondence Prof. Monika Raj, Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, United States, Email: monika.raj@emory.edu

That potential is to use AI not just as a “prediction machine” but as a *creative partner*, one that can help generate new ideas, or hypotheses, that could lead to the discovery of completely new types of reactivity, reaction mechanisms, and molecular frameworks (Fig. 1).

Breaking Cognitive Overfitting

Traditionally, synthetic organic chemistry has advanced through the cycle of observation, intuition, and experimental refinement, driven by chemist's ability to recognize subtle patterns in reactivity. But human thinking is naturally biased: our training, past experiences, and even the published literature push us toward familiar reactions. This “cognitive overfitting” means we often focus on a narrow slice of chemistry.³ AI can help counteract this by revealing reaction possibilities that fall outside our usual mental patterns.

Recent examples illustrate the point. For example, Cronin and coworkers⁴ coupled machine learning with an automated, programmable synthesis platform to explore reactivity in a closed loop. Rather than only optimizing known transformations, the system prioritized experiments that revealed unexpected outcomes, showing how self-driving laboratories can probe beyond routine reactivity. Historical antecedents exist—algorithmic approaches to discovering new reactions were reported as early as the 1990s,⁵ but modern implementations integrate continuous instrumentation, feedback control, and richer molecular representations. Similarly, the Aspuru-Guzik group has developed “inverse-design” frameworks using advanced AI models to explore chemical space in multiple directions at once, considering several target properties simultaneously, not just one.⁶ In both cases, AI acted less like a black box and more like a

collaborator offering “alien” suggestions that challenged conventional design rules.⁷

Three Shifts Needed

To make this shift more common, three changes are required. *First, broaden the training data:* AI models should learn from reaction datasets that capture diverse and underrepresented transformations. Current corpora, whether from public databases or the literature, are dominated by a few reaction classes. For example, Beker et al.⁸ reported over 10,000 cases of heteroaryl Suzuki–Miyaura couplings, highlighting the overwhelming prevalence of this single transformation. Such abundance risks models’ “learning popularity” rather than reactivity, defaulting to common reactions instead of uncovering less explored but potentially transformative chemistries. Countering this bias requires deliberately incorporating reactions such as radical cascades, photochemical rearrangements, and high-valent metal-mediated oxidations. *Second, blend AI with chemical insight:* Purely statistical models risk memorizing patterns rather than capturing underlying reactivity. Embedding mechanistic understanding—through interpretable physical–organic descriptors such as sterics, electronics, pK_a values, or kinetic parameters—allows predictions to generalize beyond the training distribution.⁹ This integration also makes outputs more chemically meaningful: a model that incorporates mechanistic context can suggest transformations that are both novel and consistent with known principles, giving chemists a stronger basis for trust and follow-up experimentation. *Third, make AI thinking visible:* Most current models present outputs as black boxes, leaving chemists unsure why a prediction was made. Interfaces should instead reveal which features or reaction motifs influenced the outcome, highlight precedent

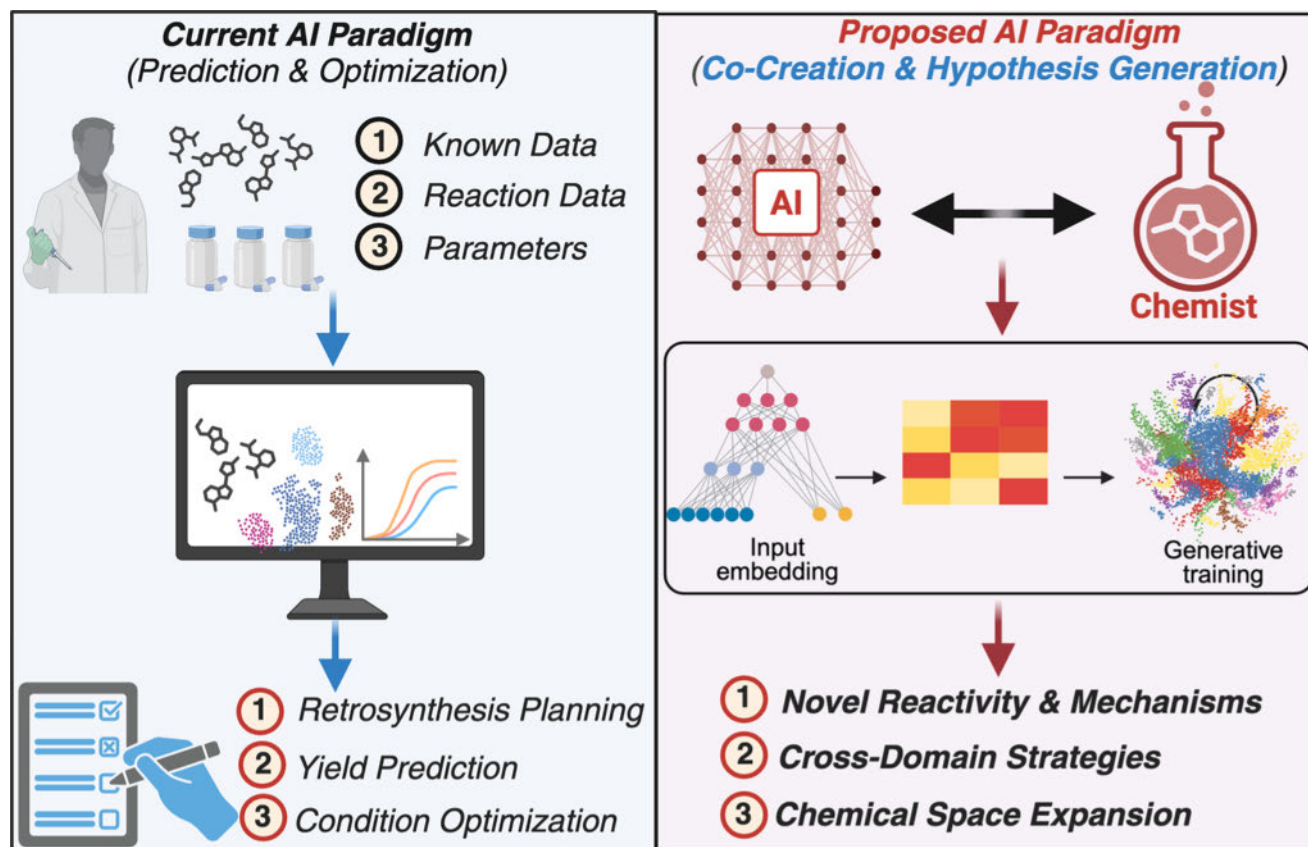


Fig. 1 From prediction to co-creation: reframing AI's role in organic synthesis. Created in BioRender. Raj, M. (2025) <https://BioRender.com/gvu05e9>

examples considered similar, and provide calibrated measures of uncertainty. Such transparency turns opaque predictions into interrogable hypotheses. This allows chemists to probe, stress-test, and refine model suggestions, transforming AI from a one-way oracle into a collaborative partner in discovery.

Beyond Small Molecules

The potential of AI extends far beyond small-molecule synthesis. In complex molecular settings such as total synthesis of natural products, diversity-oriented synthesis, or late-stage modification of bioactive scaffolds, exploring unknown chemical reactivity can be slow and expensive. AI can act as a triage tool, rapidly narrowing huge numbers of possible reactions to a manageable set of promising options. Similar opportunities exist in peptide and protein chemistry. Many bioconjugation, macrocyclization, and post-translational modification strategies are inspired by small-molecule logic. AI could go further not only improving site selectivity and orthogonality but also uncovering reaction pathways that are difficult or impossible to find by trial-and-error screening alone. Models trained on both organic and enzymatic reactions could suggest hybrid strategies bridging synthetic chemistry with biomolecular reactivity. This could speed up the creation of precise tools for chemical biology. A recent study published in *Nature*¹⁰ showed how high-throughput experimentation combined with Bayesian optimization can efficiently search reaction conditions. AI was not intended to outperform human intuition

at known chemistry but to accelerate the productive and unconventional solutions.

AI is not a final arbiter but a scientific partner—cogenerating early hypotheses, stress-testing assumptions, and enabling solutions that neither could achieve alone. For organic chemists, this means thinking of AI in a new way; not as the final step in planning the synthesis, but as a partner early in the creative process. Just as NMR spectroscopy did not replace the chemist's interpretation skills but expanded the kind of structures we could solve, AI can expand the range of synthetic ideas we can imagine.

Making this vision real will require changes to both infrastructure and mindset. We need better data standards for reactions; open databases that include failed, as well as successful experiments; and incentives for chemists to gain cross-disciplinary skills. More importantly, chemists must see AI reasoning as part of their own thinking process, not as a black box service done elsewhere. If we do this, AI will not replace the art of synthesis; rather, it will make it richer, more daring, and ultimately more creative.

Author affiliations

- 1 Department of Chemistry, Emory University, Atlanta, GA, United States.
- 2 Department of Computer Science, Emory University, Atlanta, GA, United States.

Statements and additional information

Conflict of Interest The authors declare that they have no conflict of interest.

Contributors' Statement Tuan B Vinh: Conceptualization, Writing - original draft. Tao Li: Writing - review & editing. Carl Yang: Writing - review & editing. Monika Raj: Conceptualization, Funding acquisition, Resources, Supervision, Writing - review & editing.

References

- 1 Segler MHS, Preuss M, Waller MP. *Nature* 2018; 555: 604–610
- 2 Shields BJ, Stevens J, Li J et al. *Nature* 2021; 590: 89–96
- 3 Corey EJ, Wipke WT. *Science* 1969; 166: 178–192
- 4 Granda JM, Donina L, Dragone V, Long D-L, Cronin L. *Nature* 2018; 559: 377–381
- 5 Herges R, Hooock C. *Science* 1992; 255(5045): 711–713
- 6 Sanchez-Lengeling B, Aspuru-Guzik A. *Science* 2018; 361: 360–365
- 7 Jones N. *Nature* 2025; 10.1038/d41586-025-00377-9
- 8 Beker W, Roszak R, Wołos A et al. *J Am Chem Soc* 2022; 144: 4819–4827
- 9 Reid JP, Proctor RSJ, Sigman MS, Phipps RJ. *J Am Chem Soc* 2019; 141: 19178–19185
- 10 Burger B, Maffettone PM, Gusev VV et al. *Nature* 2020; 583: 237–241