

Ask most people how artificial intelligence designs a new medicine and you tend to get the same picture: feed a computer enough data, wait for the perfect molecule, announce a cure. It makes a tidy headline. It is also a poor account of what actually happens in a medicinal chemistry lab.

What AI has proved most useful for, so far, is rather more mundane. It helps chemists invent structures worth making, tidy those structures once they exist, and — less often discussed — work out whether a promising compound can be manufactured without bankrupting the programme. None of that looks much like the science-fiction version.

The part that gets the press is the first of these: proposing candidate molecules that might bind a biological target. Generative models can now spit out thousands of chemically sensible structures with desired properties already baked in. One system, built by Insilico Medicine, produced a candidate that became the first widely reported generative-AI-designed drug to reach Phase II clinical trials, for pulmonary fibrosis. That is a real milestone. It is also easily over-read. The model proposed structures worth synthesising. Whether the resulting molecule actually worked, was safe, and could ultimately be manufactured at scale still had to be established experimentally.

Most medicinal chemists spend far more of their working lives on the next problem. Once you have a molecule, or a family of related ones, you almost never nail potency, selectivity and safety in one shot. You make small structural tweaks — fluorine for chlorine, a longer chain, a different ring — and watch what happens to activity.

This is not a new difficulty, and AI did not invent a solution to it. Chemists have been doing a version of statistical modelling since the 1960s, when Corwin Hansch showed that biological activity could often be predicted from a simple equation built on a handful of physical properties: how greasy the molecule is, how it distributes electron density, how big it is. Quantitative structure–activity relationship analysis is the direct ancestor of a great deal of what now gets called AI in chemistry. The descriptor toolbox has changed — quantum-mechanical calculations now supply numbers Hansch once measured by hand — and there is more willingness to reach for slightly more flexible tools, such as decision trees, when the relationship between structure and activity is not a straight line. The underlying job has not.

A tutorial from chemists at the University of Utah, published in *Science Advances*, puts this rather sharply, and it cuts against the idea that AI only becomes useful once the datasets are enormous. A working chemist often has only a few dozen or a few hundred data points. Each of them is the product of a real, expensive experiment. High-throughput experimentation can accelerate that gathering —

many small reactions run in parallel, rather than one flask at a time — but it does not conjure the vast labelled datasets that most public discussion of AI quietly assumes. The Utah authors argue for choosing the smallest model the problem will bear, rather than treating sparsity as a defect to be overwhelmed with parameters.

A related collaboration involving the Utah researchers makes the point in a single cut. In a nickel-catalysed cross-coupling, ninety monodentate phosphine ligands were tested. One number captured the key distinction: how much physical space each ligand occupies around the metal. Above a threshold the reaction died; below it, it worked. The descriptor also explained why. Ligands above the cut-off could not form the ligated species the nickel chemistry needed.

Then there is the question the public hears least about, and the one that regularly wrecks otherwise elegant programmes: can the molecule actually be made?

A structure that looks perfectly reasonable on paper can turn a simple-looking synthesis into weeks of frustration. A step that will not go beyond 20% yield. A chiral centre that racemises if you glance at it. An intermediate too unstable to isolate. A purification that needs three rounds of chromatography before anything usable comes off the column. Even making enough compound for the first round of in vitro testing — often just a few tens of milligrams — can swallow disproportionate time on a novel scaffold, precisely because the most interesting new molecules sit furthest from the well-trodden routes any chemist, or any AI trained on the literature, has seen before.

Retrosynthesis is supposed to help here: start from the target and work backwards to simpler starting materials and a plausible sequence of steps. Software for that job has improved a good deal. It is still only as good as the chemistry it has been shown, and a scaffold that has never been made before will send both the chemist and the model off the map.

And that is the easy end. If the compound is promising enough for an animal study, the route has to come off the milligram scale. Scaling from milligrams to grams, or grams to kilograms, is not “the same reaction, more times”. Heat no longer dissipates in the same way. Mixing is no longer instantaneous. A filtration that took thirty seconds in a flask takes six hours in a larger vessel. A reagent that was acceptable in a fume hood becomes a serious safety problem in quantity.

Economics stops being an afterthought and becomes part of the chemistry. A catalyst loading of 5 mol% is a rounding error when you are making 20 milligrams. Multiplied out to a kilogram batch, the same loading of an expensive palladium or nickel catalyst can decide whether the programme is viable. Every extra step compounds cost and eats overall yield. Every litre of solvent has to be

bought, handled and disposed of. A route elegant enough to publish is not automatically one you would want to run at scale.

This is where the same small-data methods earn their keep, only now the objective is no longer simply “highest yield”. It is highest yield with a catalyst you can afford, under conditions robust enough not to collapse on a Tuesday when the humidity is different. One case discussed in the Utah tutorial does exactly this kind of work. Process chemists developing a route to an active pharmaceutical ingredient for asthma first classified which catalysts gave usable conversion, then built a tighter model to improve selectivity, and predicted ligands that outperformed the starting conditions on both counts. Far less glamorous than generating a novel candidate from scratch. Also the difference between a molecule that reaches a patient and one that does not.

Real drug design still runs on relatively little data, hard-won at the bench, and hemmed in at every stage by the practical business of making the stuff. The methods that have lasted in this field are the ones built to live with that constraint, not the ones that pretend it away.

Glossary

Generative model. An AI system trained to invent new examples — here, new molecular structures — rather than merely classify or score ones that already exist.

Quantitative structure–activity relationship (QSAR). A mathematical model that links a molecule’s measurable physical properties to its biological activity. Most of what is now sold as “AI in chemistry” is a descendant of this.

Descriptor. A number that stands in for some physical or chemical feature of a molecule: its size, how it spreads electron density, how greasy it is, and so on. These numbers are what the statistical model actually sees.

Decision tree. A deliberately simple model. It sorts data by asking a run of yes/no questions about descriptor values, and ends with a prediction.

Ligand. A small molecule bound to a metal catalyst. It governs how the metal reacts; change the ligand and a reaction that worked can stop working.

Buried volume. How much physical space a ligand occupies around the metal centre it is attached to.

Catalyst loading. How much catalyst you use relative to the reactants, usually given as mole percent (mol%). At laboratory scale this is often a rounding error; at manufacturing scale it is a cost.

Chiral centre / racemise. A chiral centre is a structural feature that can give a molecule a particular three-dimensional “handedness”. To racemise is to lose that handedness, producing a mixture of the two forms.

Retrosynthesis. Working backwards from the target molecule to simpler starting materials, and to the sequence of steps that would build it.

Process chemistry. The job of turning a synthesis that works in a flask into one that can be run reliably, safely and cheaply at a larger scale.

High-throughput experimentation (HTE). Running many small-scale reactions in parallel so that experimental data arrive quickly rather than one flask at a time.

In vitro / in vivo. Literally “in glass” and “in the living”: a dish or test tube, versus an animal or a human.

Further reading

Haas, B. C., Kalyani, D. & Sigman, M. S. Applying statistical modeling strategies to sparse datasets in synthetic chemistry. *Science Advances* 11, eadt3013 (2025). <https://doi.org/10.1126/sciadv.adt3013>

Newman-Stonebraker, S. H. et al. Univariate classification of phosphine ligation state and reactivity in cross-coupling catalysis. *Science* 374, 301–308 (2021). <https://doi.org/10.1126/science.abj4213>

Dotson, J. J. et al. Data-driven multi-objective optimization tactics for catalytic asymmetric reactions using bisphosphine ligands. *Journal of the American Chemical Society* 145, 110–121 (2023). <https://doi.org/10.1021/jacs.2c08513>

Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* 360, 186–190 (2018). <https://doi.org/10.1126/science.aar5169>

Chuang, K. V. & Keiser, M. J. Comment on “Predicting reaction performance in C–N cross-coupling using machine learning”. *Science* 362, eaat8603 (2018). <https://doi.org/10.1126/science.aat8603>

Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 555, 604–610 (2018). <https://doi.org/10.1038/nature25978>

Hansch, C. & Fujita, T. ρ - σ - π Analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society* 86, 1616–1626 (1964). <https://doi.org/10.1021/ja01062a035>



Insilico Medicine. First generative AI drug begins Phase II trials with patients.
https://insilico.com/blog/first_phase2