

# AI learns the language of chemistry to predict how to make medicines

September 3 2019

---



Credit: Денис Марчук from Pixabay

Researchers have designed a machine learning algorithm that predicts the outcome of chemical reactions with much higher accuracy than trained chemists and suggests ways to make complex molecules, removing a significant hurdle in drug discovery.

University of Cambridge researchers have shown that an algorithm can predict the outcomes of complex [chemical](#) reactions with over 90%

accuracy, outperforming trained chemists. The algorithm also shows chemists how to make target compounds, providing the chemical "map" to the desired destination. The results are reported in two studies in the journals *ACS Central Science* and *Chemical Communications*.

A central challenge in drug discovery and materials science is finding ways to make complicated organic molecules by chemically joining together simpler building blocks. The problem is that those building blocks often react in unexpected ways.

"Making molecules is often described as an art realized with trial-and-error experimentation because our understanding of chemical reactivity is far from complete," said Dr. Alpha Lee from Cambridge's Cavendish Laboratory, who led the studies. "Machine learning algorithms can have a better understanding of chemistry because they distill patterns of reactivity from millions of published chemical reactions, something that a [chemist](#) cannot do."

The algorithm developed by Lee and his group uses tools in [pattern recognition](#) to recognize how chemical groups in molecules react, by training the [model](#) on millions of reactions published in patents.

The researchers looked at chemical reaction prediction as a machine translation problem. The reacting molecules are considered as one language, while the product is considered as a different language. The model then uses the patterns in the text to learn how to translate between the two languages.

Using this approach, the model achieves 90% accuracy in predicting the correct product of unseen [chemical reactions](#), whereas the accuracy of trained human chemists is around 80%. The researchers say that the model is accurate enough to detect errors in the data and correctly predict a plethora of difficult reactions.

The model also knows what it doesn't know. It produces an uncertainty score, which eliminates incorrect predictions with 89% accuracy. As experiments are time-consuming, accurate prediction is crucial to avoid pursuing expensive experimental pathways that eventually end in failure.

In the second study, Lee and his group, collaborating with the biopharmaceutical company Pfizer, demonstrated the practical potential of the method in drug discovery.

The researchers showed that when trained on published chemistry research, the model can make accurate predictions of reactions based on lab notebooks, showing that the model has learned the rules of chemistry and can apply it to drug discovery settings.

The team also showed that the model can predict sequences of reactions that would lead to a desired product. They applied this methodology to diverse drug-like molecules, showing that the steps that it predicts are chemically reasonable. This technology can significantly reduce the time of preclinical drug discovery because it provides medicinal chemists with a blueprint of where to begin.

"Our platform is like a GPS for chemistry," said Lee, who is also a Research Fellow at St Catharine's College. "It informs chemists whether a reaction is a go or a no-go, and how to navigate reaction routes to make a new molecule."

The Cambridge researchers are currently using this reaction prediction technology to develop a complete platform that bridges the design-make-test cycle in [drug discovery](#) and materials discovery: predicting promising bioactive molecules, ways to make those complex organic [molecules](#), and selecting the experiments that are the most informative. The researchers are now working on extracting chemical insights from the model, attempting to understand what it has learned that humans

have not.

"We can potentially make a lot of progress in chemistry if we learn what kinds of patterns the model is looking at to make a prediction," said Peter Bolgar, a Ph.D. student in synthetic organic chemistry involved in both studies. "The model and human chemists together would become extremely powerful in designing experiments, more than each would be without the other."

**More information:** Philippe Schwaller et al. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction, *ACS Central Science* (2019). [DOI: 10.1021/acscentsci.9b00576](https://doi.org/10.1021/acscentsci.9b00576)

Alpha Lee et al. Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space, *Chemical Communications* (2019). [DOI: 10.1039/C9CC05122H](https://doi.org/10.1039/C9CC05122H)

Provided by University of Cambridge

Citation: AI learns the language of chemistry to predict how to make medicines (2019, September 3) retrieved 1 May 2024 from <https://techxplore.com/news/2019-09-ai-language-chemistry-medicines.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.