Editorial

Rates against the machine

Check for updates

Computational chemistry has become an increasingly common part of catalysis research. More recently, data-based methods such as machine learning have been suggested as a means to speed up discovery. This Focus issue features a collection of content dedicated to machine learning as pertaining to its potential impact on the field of catalysis.

The term artificial intelligence (AI) seems to have moved from specialist to everyday vernacular in recent years. We encounter it on the news, on social media, on our smart devices. A recent controversy brought AI to the forefront in the publishing industry when it emerged that some articles had been published crediting ChatGPT — an AI-based chatbot that can generate human-like responses — as an author[1]. While consensus was rapidly reached that ChatGPT did not meet the criteria for the definition of authorship, ethical considerations regarding whether ChatGPT should be used, how it should be acknowledged, and if it should be regarded as plagiarism remain under debate.

ChatGPT makes use of machine learning (ML), which is defined as "the capacity of computers to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and infer from patterns in data"[2]. ML is also increasingly making its presence felt in scientific research, including and most relevant to our readership, catalysis science. While the delegate list of any recent catalysis conference reveals the strong representation of theoretical chemists amongst the community, the majority of theory attached to catalysis papers to date has been based on more conventional methods, such as density functional theory (DFT).

In a typical study, an atomic model of a catalyst would be constructed based on the available characterization, and the energies of the interaction between catalyst, reagents, intermediates and products are calculated. The problem with most catalytic systems is that they are enormously complex. Even if the precise structure of the catalyst can be

determined — a big if in itself — the conditions, solvent and numerous other variables need to be taken into account. Hence while theory has been able to provide valuable insights into individual systems, the high computational demand has restricted the ability to use theoretical methods to rapidly speed up the prediction and design of successful catalysts. It is precisely towards this issue that ML may be able to offer some solutions because it does not rely on a total understanding of the existing catalytic systems.

With this in mind, this issue of *Nature Catalysis* focuses on current trends in ML in the context of catalysis science, featuring a selection of articles which seek to draw these communities together in pursuit of progress. In their Review, Hongliang Xin and colleagues discuss the various ways in which ML can assist in closing the gap between highly involved catalytic systems and what can be readily computed at scale, in order to facilitate more rapid advances in catalyst discovery.

If one had to choose a single word to describe a real catalytic system, then complex is probably the right one, due to the intricacies of reaction mechanisms. And yet detailed knowledge of the path that connects reactant with products is essential to advance catalysis. In this spirit, Johannes Margraf, Karsten Reuter and co-workers present a Perspective on how ML can aid the exploration of complex reaction networks and group the approaches into bottom-up and top-down exploration strategies. The former includes, for instance, the use of surrogate models to fit DFT energies — a

very popular strategy — whereas the latter examines how ML can be combined with experimental data to derive microkinetic and mechanistic details.

> **"It is precisely towards this issue that machine learning may be able to offer some solutions because it does not rely on a total understanding of the existing catalytic systems."**

Another area where ML is applied to complex reaction networks is the design of enzyme cascades. In their Review, Huimin Zhao and colleagues discuss the impact of ML in synthesis planning, enzyme identification and engineering, and pathway optimization for improved retrobiosynthesis of molecules in vitro and in vivo.

Yet, no matter how smart a machine can be programmed, the extent to which it can learn will always be limited by the sufficiency of the data with which it is fed. Data quality, quantity and bias are key issues for the data science community when attempting to solve any ML problem. In their Comment, Toshiaki Taniike and Keisuke Takahashi discuss the effects of data bias — particularly the lack of reporting of negative results or other data not seen as useful by experimental catalysis researchers — on efforts in ML. The intention is that in raising awareness of the issue, the catalysis and data

Editorial

science communities will be able to better work together to achieve mutual progress.

If one thing is clear, the era of AI and ML is only just beginning. It is our sincere belief that the catalysis community will embrace all available tools, including ML, to bring forward solutions to many of the world's most pressing problems. We hope that you find inspiration within these pages for humankind to perfect its use of machines so that perhaps, one day, rates — and many other catalytic properties — may become predictable with a high degree of confidence, emerging against a background that is currently dominated by trial and error.

References

1. Stokel-Walker, C. *Nature* **613**, 620–621 (2023).
2. Machine learning. *Oxford English Dictionary* (Accessed 13 Feb 2022); https://www.oed.com/view/Entry/111850?redirectedFrom=machine+learning#eid38479194