



READER

SOMMERAKADEMIE 2026

TRANSFORMATION

PRO **SCIENTIA**
STUDIENSTIFTUNG

READER

PRO SCIENTIA Sommerakademie 2026

TRANSFORMATION

Künstliche Intelligenz in der Chemie

MAXIMILIAN PETER-PAUL KOVAR

Wien

Während Large Language Models (LLMs) die öffentliche Diskussion über maschinelles Lernen (ML) dominieren, revolutionieren in den Naturwissenschaften andere ML-Modelle unsere Welt. Dieser Text schildert, wieso dem so ist, erklärt dabei grundlegende Konzepte des noch jungen Wissenschaftsfeldes und geht weiters auf die Chancen, die sich, insbesondere in der Chemie, daraus ergeben, ein. Ein besonderer Fokus des Textes liegt hierbei auf der Vorhersage chemischer Reaktionseigenschaften.

Spätestens nach dem Erscheinen von OpenAIs Sprachmodell 'ChatGPT' ist eines klar: Keine Disziplin der Natur-, Geistes- oder Sozialwissenschaften wird von der Technologie des maschinellen Lernens – umgangssprachlich oft mit 'künstlicher Intelligenz' (KI) gleichgesetzt – gänzlich unberührt bleiben. Vor allem Sprachmodelle (*large language models*, LLMs) haben jetzt schon Folgen, die in allen wissenschaftlichen Disziplinen spürbar sind: Sie verändern unsere Recherchepraxis, die Art, wie Texte entstehen, und stellen Lehre und Prüfungskultur infrage. In allen Fächern wird es also ein 'vor der KI' geben, so wie es ein 'vor dem Internet' und ein 'vor dem Computer' gab. In vielen Disziplinen sind es aber nicht nur LLMs, die bleibende Spuren hinterlassen werden. In manchen werden sie sogar nur eine Fußnote des Fortschrittes bedeuten, den maschinelles Lernen ermöglicht.

Dieser Artikel erklärt zunächst, welche einfache Mathematik sich hinter vielen Verfahren des maschinellen Lernens verbirgt. Darauf aufbauend wird erläutert, was von LLMs erwartet werden darf bzw. muss und was nicht. Anschließend wird begründet, warum diese Modelle in Disziplinen wie der Chemie nur eine untergeordnete Rolle spielen, wobei es notwendig sein wird, die ungefähre Vorgehensweise chemischer Forschung und insbesondere die Rolle der chemischen Intuition zu beleuchten. Schließlich wird sich der Text eigenen Erfahrungen und Forschungsfragen widmen, bevor er mit einer allgemeinen Einschätzung schließt, die Hoffnung

macht, zukünftig weniger Sackgassen in der Forschung betreten zu müssen.

Wie mathematische Neuronen lernen

Im Kern handelt es sich bei maschinelltem Lernen um eine vergleichsweise simple Familie statistischer Verfahren, die als Regressions- (oder Klassifikations-) analyse bezeichnet werden. Eine der wichtigsten Annahmen hinsichtlich eines zu lösenden Problems ist demnach zuallererst jene, dass es durch eine mathematische Funktion, die ein Ergebnis (*Output*) auf Basis von Variablen (*Input*) berechnet, beschreibbar ist.

$$f: (\text{Variablen}) \mapsto \text{Ergebnis}$$

Dies kann zum Beispiel die Frage nach einer chemischen Reaktionseigenschaft sein, welche von den Molekülen abhängt, die reagieren; oder ein Buchstabe in ebendiesem Satz, der von all den Buchstaben abhängt, die vor ihm kamen.

$$f: (\text{Moleküle}) \mapsto \text{Reaktionseigenschaft}$$

$$f: (\text{Buchstaben}) \mapsto \text{Buchstabe}$$

Beide hier genannten Beispiele haben eines gemeinsam: Diese Funktionen müssen eine große Menge an Variablen beinhalten, die nicht auf simple Weise (zum Beispiel linear) miteinander oder mit dem Ergebnis verknüpft sind. Man kann also nicht, wie in einer linearen oder polynomiellen Regressionsanalyse, von einer bestimmten Form (linear, polynomiell, ...) der Funktion ausgehen. Daher wählt man eine Klasse sehr flexibler Funktionen mit vielen einstellbaren Parametern: *Neuronale Netze*. Diese sind nämlich *universelle Funktionsapproximatoren*. Das bedeutet, sie können prinzipiell jede beliebige kontinuierliche Funktion nachbilden, ohne dass deren genaue Form im Voraus bekannt sein muss. Möglich ist dies durch eine große Zahl an *Neuronen*, die – für sich genommen – allerdings simple mathematische

Objekte sind. Jedes von ihnen berechnet eine gewichtete Summe seiner *Inputs*, addiert einen Wert und transformiert das Ergebnis mittels einer sogenannten Aktivierungsfunktion (deren genaue Form für diesen Beitrag nicht relevant ist). Viele solcher Neuronen, in Schichten (*layers*) angeordnet, bilden zusammen das Netz, das in seiner Gesamtheit immer noch eine Funktion darstellt. Die Gewichte (*weights*) und addierten Werte (*biases*) werden zusammen als *Parameter* bezeichnet und sind die eigentliche Information, über welche das Neuronale Netz verfügt. Die Form der gesamten Funktion wird einfach nach und nach durch das iterative Verändern der Parameter angepasst, was als *training*, umgangssprachlich als 'Lernen' bezeichnet wird und den Unterschied zu anderen Formen der Regressionsanalyse darstellt. Die Basis dafür bildet eine meist sehr große Menge beobachteter *Input-Output*-Paare, die möglichst gut reproduziert werden sollen. Dazu definiert man eine *loss function*, die quantifiziert, wie weit die Vorhersage des Modells vom tatsächlichen Ergebnis abweicht. Zunächst wird das Modell mehr oder weniger zufällige Vorhersagen treffen, die *loss function* wird also manchmal groß, manchmal klein sein. Der Schlüssel zum Lernen liegt allerdings nicht im Wert, den sie für Bestimmte Inputs annimmt, sondern in ihrer Änderung bezüglich der einzelnen Parameter des Neuronalen Netzes, die angepasst werden sollen. Mathematisch gesehen geht es also um ein Differenzial – genauer gesagt den Gradienten bzw. das totale Differenzial – dessen Bildung in solch einem Fall immer möglich ist, da es sich um eine Kombination von wiederum ableitbaren Funktionen (jenen der Neuronen) handelt. Im Englischen spricht man hier von *back-propagation*. Auf Basis dessen werden alle Parameter geringfügig korrigiert: jene die den Fehler stark vergrößern, mehr; solche, die geringen Einfluss auf ihn haben, weniger. Durch wiederholtes Durchlaufen der Trainingsdaten und schrittweises Nachjustieren der Parameter 'lernt' das Modell, die *Input-Output*-Paare immer besser zu reproduzieren.

Bis hierher funktioniert das beschriebene Prinzip für beide genannten Beispiele gleichermaßen. Ob Buchstaben oder Moleküle: Der Lernalgorithmus ist im Grunde derselbe. Der entscheidende Unterschied liegt in der Natur der Daten und darin, was als 'korrekte Vorhersage' gilt. Bei der Vorhersage von Reaktionsenergien ist das Ergebnis eine messbare, physikalisch definierte Größe. Bei einer Buchstabenvorhersage hingegen geht

es um bedingte Wahrscheinlichkeiten: Will man den nächsten Buchstaben ohne jeglichen Kontext vorhersagen, genügt eine Verteilung über das Alphabet (26 Möglichkeiten unter Ausschluss von Majuskeln und sonstigen Zeichen). Kennt man den vorhergehenden Buchstaben, wird daraus eine bedingte Wahrscheinlichkeit – aus 26 Werten werden 26×26 Kombinationen. Bei Berücksichtigung von zwei vorhergehenden Buchstaben (als sogenannter 'Kontext') sind es bereits $26^2 \times 26$, und so weiter. Alle Wahrscheinlichkeiten vorab zu berechnen und im Speicher des Computers zu behalten um daraus iterativ einen nächsten Buchstaben zu 'erwürfeln' ist nicht praktikabel, wenn die generierte Sprache über ganze Wörter und Sätze konsistent wirken soll. Damit Sprachmodelle keine gigantischen Speicher für Wertetabellen benötigen, müssen sie also eine Funktion lernen. Diese muss aber von immer längeren Kontexten abhängen.

Computer lernen Sprache

Bevor wir uns in diesem Text aber Gedanken über immer längere Kontexte machen können, stoßen wir aber zunächst auf das fundamentale Problem der Repräsentation. Computer arbeiten nicht mit Sprache im menschlichen Sinne, sondern mit Zahlen. Dieser Text, den ich gerade tippe, ist für den Computer eine lange Sequenz von Zahlen. Weiters trägt die alphabetische Reihenfolge der Buchstaben im Alphabet keine semantische Bedeutung. Buchstaben – wie im Beispiel zuvor – also nur als eine Zahl des Alphabets wiederzugeben, ermöglicht es einem Computer nicht, die größere Ähnlichkeit von "a" zu "o" im Vergleich zu "a" zu "q" zu sehen. Lässt man dem Modell jedoch die Freiheit, diese Zahlenreihen selbstständig zu interpretieren und als Vektoren abzubilden – eine Technik, die man als *embedding* bezeichnet –, offenbart sich ein tieferes Muster: Buchstaben mit ähnlicher Rolle oder phonetischer Ähnlichkeit weisen in einem solchen Vektorraum auch räumlich ähnliche Positionen auf.

Um nun aber immer größere Kontexte für die jeweils nächste Vorhersage in Betracht ziehen zu können, ist es sinnvoll, die Idee der einzelnen Buchstaben ganz hinter sich zu lassen. Dies geschieht durch sogenannte *Tokenisierung*. Sprache wird hierbei in kleine Einheiten zerlegt, sogenannte Tokens. Zu einem Token fasst man jene Buchstabenfolgen zusammen, die besonders häufig im

gesamten Trainingscorpus vorkommen und gibt ihnen eine eigene Zahl, mit der der Computer diesen Token repräsentiert. Das Wort "Strawberry" ist für das LLM *gpt-4o* also nicht "19-20-18-1-23-2-5-18-18-25" (S-T-R-A-W-B-E-R-R-Y), sondern "3504-1134-19772" (S-T-R-A-W-B-E-R-R-Y), weswegen es nur aus 3 Zahlen besteht. Dies ermöglicht es, mehr Kontext für die Voraussage des nächsten Tokens zu berücksichtigen, führt aber auch zu unerwünschten Nebenwirkungen. Fragt man nun nämlich das LLM "How many 'R's are there in 'Strawberry'?" (in Tokens: "5299-1991-461-49-885-553-1354-306-461-3504-1134-19772-127222"), ist "three" (Token: "47660") nur zufällig die richtige Antwort (wenn überhaupt). Obwohl Sprachmodelle heutzutage in der Regel sogenannte *transformer* sind, deren Architektur dem neuronalen Netz ermöglicht zu lernen, welche Tokens des Kontexts für den nächsten relevant sind, führt dies lediglich zu Ergebnissen, die zwar im besten Fall von menschlicher Sprache ununterscheidbar sind – nicht aber zu tatsächlichem Verständnis (Bender et al. 2021).

Hier liegt eine der Grenzen des maschinellen Lernens im Sprachkontext. LLMs sind hochkomplexe statistische Modelle, aber sie besitzen kein Verständnis von Semantik oder Logik. Sie wissen nicht, was eine "Strawberry" ist, sie kennen nur die statistische Wahrscheinlichkeit, welche Tokens in welcher Reihenfolge aufeinander folgen. Auf Basis von Milliarden von Texten ist es wahrscheinlich (oder auch nicht), dass es zu einer richtigen Antwort kommt. Kein deterministischer, sondern ein stochastischer Algorithmus führt dabei zum Ergebnis.

Daraus folgt eine fatale Konsequenz für ihren Einsatz in der wissenschaftlichen Arbeit: Man kann ein LLM nicht einfach umprogrammieren. Man kann die zugrunde liegende Wahrscheinlichkeitsverteilung nicht direkt beeinflussen, ohne das gesamte Modell neu zu trainieren. 100%-ige Korrektheit ist daher nie garantierbar. LLMs werden immer halluzinieren – also Dinge behaupten, die statistisch plausibel klingen, aber faktisch falsch sind. Man könnte auch sagen, sie halluzinieren und haben lediglich meistens Glück, dass ihre Aneinanderreihungen von Tokens oft wahr sind.

Während LLMs mithin die Welt in allen Bereichen verändern und nur Menschen letztlich dafür die Verantwortung tragen können, ob dies eine Veränderung zum Guten sein wird, gibt es in vielen Disziplinen, so auch in

meiner, der Chemie, noch weit mehr Anwendungen maschinellen Lernens, die von Grund auf anders funktionieren.

Die Sprache der Natur

Menschliche Sprache ist allerdings – selbst im akademischen Kontext – oft uneindeutig oder ungenau und somit als Repräsentation der Realität für Vorhersagen in den Naturwissenschaften suboptimal. Ein und dasselbe Molekül kann Dutzende Namen haben. Aber wie bei der Tokenisierung des Wortes "Strawberry" sieht ein LLM "Eisessig", "Ethansäure" und "O=C(O)C" als völlig verschieden, wobei es sich aber jeweils nur um den Trivialnamen, den Systematischen Namen und den SMILES-String von Essigsäure handelt, deren Eigenschaften stets die gleichen bleiben. Umgekehrt können ähnliche Namen unterschiedliche Strukturen bezeichnen: Sulfat und Sulfit unterscheiden sich nur durch einen Buchstaben, beschreiben aber gänzlich verschiedene Verbindungen.

Diese Schwachstelle bei der Verwendung von LLMs in der Chemie zeigt sich noch prononcierter, wenn es um die exakte räumliche Anordnung von Atomen – die Stereochemie – geht, entscheidet sie doch oft über Wirkung oder Unwirksamkeit, über Heilung oder Toxizität, wie spätestens seit der Contergan-Katastrophe bekannt sein dürfte. Ein Sprachmodell, das nur Text verarbeitet, kann diese Information nicht erfassen, es sei denn, sie wird explizit beschrieben. Doch selbst dann bleibt fraglich, ob die Beschreibung dem Problem gerecht wird. Gewiss verändern auch in der Chemie LLMs die Art und Weise, wie geforscht wird. Aber die eigentliche Revolution liegt nicht in jenen Modellen, die Sprache generieren, die Moleküle beschreiben, sondern in solchen, die direkt lernen, wie Moleküle agieren und reagieren. Im Folgenden geht es daher nicht um LLMs, sondern um eine andere Form von maschinellen Lernen: Werkzeuge zur Vorhersage chemischer Eigenschaften und chemischer Reaktivität.

Dass die in diesem Feld bereits eine Transformation begonnen hat, zeigt die Vergabe einer Hälfte des Nobelpreises 2024 an Demis Hassabis und John M. Jumper "für die Vorhersage von Proteinstrukturen". Dieser Würdigung liegt das Programm "AlphaFold" (Jumper et al. 2021) zugrunde, das 2018 und 2020 veröffentlicht wur-

de – also nur sechs beziehungsweise vier Jahre vor der Preisvergabe. In der Welt der Nobelpreise, in der Auszeichnungen oft Jahrzehnte nach den zugrunde liegenden Arbeiten verliehen werden, ist das außergewöhnlich.

Was AlphaFold leistet, lässt sich so erklären: Aus der linearen Abfolge der Aminosäuren eines Proteins sagt das Modell mit hoher Genauigkeit dessen drei-dimensionale Struktur voraus – also jene Form, in der das Protein tatsächlich im Körper wirkt. Damit wurde eine Aufgabe, die zuvor Jahre an Experimenten erfordern konnte, zu einem Rechenproblem, das sich mit Supercomputern in Stunden lösen lässt. AlphaFold steht prototypisch für das, was maschinelles Lernen in der Chemie leisten kann: Sie ersetzt keine Theorie und kein Experiment, eröffnet aber Wege, die zuvor praktisch unzugänglich waren.

AlphaFold ist kein Einzelfall: Löslichkeiten, Toxizitäten, Bindungsaffinitäten und viele andere Moleküleigenschaften werden bereits heute routinemäßig mit maschinell gelernten Modellen abgeschätzt, um zum Beispiel in der Pharmabranche aus Millionen denkbarer Strukturen einige wenige vielversprechende Kandidaten zu wählen, die dann tatsächlich synthetisiert und experimentell geprüft werden. So kommt man schneller und kostengünstiger als noch vor ein paar Jahren zu neuen Arzneimitteln.

Doch Proteinstrukturen sind nur der Anfang. Die eigentliche Herausforderung der Chemie liegt in der Vorhersage chemischer Reaktionen (Meuwly 2021). Jeder elementare Reaktionsschritt wird im Wesentlichen durch zwei Kenngrößen bestimmt: die Reaktionsenergie (ΔG_r) und die Aktivierungsenergie ($\Delta G^\ddagger$). Die Reaktionsenergie sagt aus, ob eine Reaktion thermodynamisch überhaupt möglich ist – also ob die Produkte stabiler sind als die Edukte, aus denen erstere hervorgegangen sind. Die Aktivierungsenergie bestimmt, vereinfacht gesagt, wie schnell die Reaktion abläuft – also ob sie in praktikabler Zeit stattfindet oder Jahrmillionen braucht.

Kennt man diese beiden Größen, kann man vorhersagen, ob eine Syntheseroute überhaupt funktionieren kann, bevor auch nur ein Reagenzglas gefüllt wird. Genau hier liegt das Potenzial maschinellen Lernens: Würde ein Modell diese Kenngrößen in Sekundenbruchteilen mit hinreichender Genauigkeit vorhersagen – ähnlich wie ein LLM den nächsten Token – könnten meine

Kolleg:innen im Labor unmögliche Routen sofort ausschließen. Statt monatelang Sackgassen zu verfolgen, wüssten sie vorab, welche Wege vielversprechend sind. Das Ziel ist nicht, Experimentator:innen zu ersetzen, sondern ihnen eine:n Assistent:in an die Seite zu stellen (Szymkuć et al. 2016; Back et al. 2024), so wie ihnen LLMs beim Verfassen von Texten helfen können – nur mit dem entscheidenden Unterschied, dass hier physikalisch fundierte Größen vorhergesagt werden, nicht statistisch plausible Wortfolgen.

Dass Sprache weit entfernt von einer idealen Repräsentation der (chemischen) Wirklichkeit ist, wurde bereits erläutert. Doch was wäre die ideale Repräsentation? Kurz gesagt: Es gibt sie nicht. In den Wissenschaften gilt seit jeher: Man verwende das einfachste Modell, das die Forschungsfrage hinreichend beantwortet, nicht das fundamentalste. Jede chemische Reaktion lässt sich – genauso wie die Funktion einer Zelle oder das morgige Wetter – in der Theorie auf wenige physikalische Prinzipien zurückführen: die Quantenmechanik. Warum also nicht alle Naturwissenschaft (und in weiterer Folge vielleicht auch Sozial- und Wirtschaftswissenschaft) den Quantenphysiker:innen überlassen? Praktisch braucht es dennoch Chemiker:innen, Biolog:innen und Meteorolog:innen (und auch noch einige Wissenschaften mehr). Nicht, weil die Prinzipien der Physik falsch wären, sondern weil die direkte Berechnung aus ersten Prinzipien für komplexe Systeme oft unpraktikabel, aber öfter noch gar nicht nötig ist. Die Wellenfunktion eines Proteins exakt zu berechnen, übersteigt selbst die Kapazitäten modernster Supercomputer. Doch wie AlphaFold zeigt, muss man die Wellenfunktion nicht kennen, um die Struktur vorherzusagen. Ein Modell muss hinreichend gut sein, nicht vollumfänglich.

Bereits vor AlphaFold erlaubte es Molekulardynamik – ein Teilbereich der Chemie, der mit klassischen (also nicht quantenmechanischen) Näherungen arbeitet – Struktur und Funktion von Proteinen vorherzusagen, ohne dass ihre Wellenfunktion berücksichtigt wurde. Auch chemische Reaktionen können oft ohne jegliche Kenntnis einer errechneten Wellenfunktion vorhergesagt werden. Genau hier kommt die sogenannte ‚chemische Intuition‘ ins Spiel, deren Ausbildung bis heute ein zentraler Bestandteil aller Curricula ist. Doch was verbirgt sich eigentlich hinter diesem esoterisch klingenden Begriff?

Künstliche Intuition

Chemische Intuition ist kein Bauchgefühl, sondern eine wissenschaftlich valide Form des Mustererkennens. Sie wird über Jahre des Studiums und der Forschungspraxis antrainiert – durch das Lösen tausender Synthesaufgaben, das Lesen hunderter Publikationen und das eigene Experimentieren im Labor. Eine erfahrene Chemikerin kann abschätzen, ob eine Reaktion funktionieren wird, welche Nebenprodukte entstehen könnten und welches Lösungsmittel geeignet ist. Sie tut dies, ohne jede Möglichkeit einzeln exakt zu berechnen.

Was geschieht kognitiv, wenn eine Chemikerin eine solche Einschätzung trifft? Sie betrachtet die Molekülstrukturen von Reaktanten und Produkten, erkennt funktionelle Gruppen, bewertet elektronische Effekte, berücksichtigt sterische Hinderung, denkt an ähnliche Reaktionen aus der Literatur und gewichtet all diese Faktoren gegeneinander. Sie denkt dabei in Kategorien wie 'sauer' und 'basisch', 'nukleo-' und 'elektrophil', sterisch 'zugänglich' und 'gehindert', chemisch 'hart' und 'weich', 'elektronenarm' und '-reich'. Mathematisch gesprochen verarbeitet sie eine hochdimensionale Eingabe (die Strukturmerkmale der Edukte und möglicher Produkte) durch eine komplexe, nichtlineare Funktion und erhält eine Ausgabe (die Einschätzung der Reaktionswahrscheinlichkeit).

$f: (\text{Basizität, Nukleophilie, Sterik, ...}) \mapsto \text{Reaktionseigenschaft}$

Die einzelnen Variablen – Elektronegativität, Sterik, Basizität, Nukleophilie, chemische Härte – sind dabei zahlreich und untereinander keineswegs unabhängig. Eine erhöhte Elektronendichte an einem Stickstoffatom macht dieses nicht nur nukleophiler, sondern auch basischer. Diese Korrelationen sind nicht linear und oft kontextabhängig. Genau diese Art von Problem – viele nicht-orthogonale Parameter, zu komplex für eine geschlossene Gleichung, aber mit systematischen Trends – ist die Domäne maschineller Lernverfahren.

Chemische Intuition wirkt also bei genauerer Betrachtung wie ein hervorragendes maschinelles Lernproblem, das lediglich mangels technischer Rahmenbedingungen bisher in menschlichen Gehirnen gelöst wurde. Die Frage ist nicht, ob Maschinen das lernen können, sondern wie wir sie trainieren müssen, um es ebenso gut oder besser (zum Beispiel durch das Beziffern der Energieunterschiede) zu tun.

Hier liegt jedoch das eigentliche Dilemma. Ein mensch-

liches Gehirn lernt chemische Intuition durch sorgfältig kuratierte Erfahrungen: Lehrbuchbeispiele, Laborpraktika, betreute Forschung. Die Daten sind qualitativ hochwertig, mechanistisch erklärt und didaktisch aufbereitet. Maschinelle Modelle hingegen werden aktuell oft mit großen, unüberprüften Datensätzen trainiert. Diese basieren meist auf häufig in der Literatur oder in Datenbanken vorkommenden Reaktionen oder werden aus Molekülen, deren Bindungen dann systematisch abgeändert werden, generiert (De Landsheere et al. 2026). Dadurch sind bereits bekannte, erfolgreiche Reaktionen überrepräsentiert, Fehlschläge fehlen aber gänzlich; komplexere Überlagerungen von Mustern sind enthalten, einzelne Lehrbuchbeispiele aber nicht.

Meine eigene Forschung setzt genau an diesem Punkt an. Denn anders als bei Sprache bewegen wir uns in der Chemie in einem sogenannten *low-data regime*, haben also nur einen winzigen Teil aller möglichen Datenpunkte zur Verfügung. Quantenmechanische Berechnungen sind rechenintensiv (ganze Tage pro Reaktion auf Supercomputern), Experimente im Labor meist noch ressourcenaufwändiger. Die Auswahl der Daten und die Reihenfolge, in der sie zum Training herangezogen werden, haben daher höchstwahrscheinlich einen großen Einfluss auf die Vorhersageleistung. Wenn chemische Reaktionen systematischen Trends folgen, sollte ein Modell diese effizient lernen können, sofern die Trainingsdaten sorgfältig gestaltet sind (Strieth-Kalthoff et al. 2024).

Das führt zu mehreren konkreten Forschungsfragen: Kann ein Modell, das mit chemisch kuratierten, mechanistisch validen Daten trainiert wurde, größere, aber weniger aussagekräftige Datensätze übertreffen? Lässt sich durch *curriculum learning* (Bengio et al. 2009) – eine didaktisch sinnvolle Reihenfolge der Trainingsbeispiele – die Anzahl der benötigten Beispiele reduzieren? Und kann *active learning* identifizieren, welche Reaktionen als nächstes berechnet oder experimentiert werden sollen? Dabei wird die Nützlichkeit jedes Datenpunktes anhand der bereits bekannten und der Vorhersageunsicherheit verschieden trainierter Modelle abgeschätzt (Smith et al. 2018; Heid et al. 2023). So lassen sich die teuersten Berechnungen und Experimente auf das unvermeidbare Minimum reduzieren. Allerdings müssen wir Forschenden dabei sicherstellen, dass diese Werkzeuge mit chemischem Verständnis entwickelt werden, nicht trotz ihm.

Was sich ändern wird

Die eigentliche Revolution des maschinellen Lernens in der Chemie liegt also nicht darin, menschliche Intuition zu ersetzen, sondern sie zu systematisieren und zu skalieren. Ein gelerntes Modell in unserem Feld ist keine Black Box, der wir blind vertrauen müssen. Es ist ein Werkzeug, dessen Training wir mit chemischem Verständnis gestalten können und sollten.

Ein solches Modell, das Reaktions- und Aktivierungsenergien in Sekunden vorhersagt, ist das Werkzeug, das es Experimentator:innen erlaubt, Syntheserouten am Computer zu screenen, bevor Ressourcen in Laborversuche fließen. Das spart nicht nur Zeit und Geld – es ermöglicht Forschenden, ihre Intuition für die wirklich interessanten Fragen zu nutzen, statt sie in *trial-and-error* zu verschwenden. Es ist kein Ersatz für menschliche Expertise, sondern hilft lediglich Sackgassen als solche zu erkennen, noch bevor sie beschritten werden.

Doch maschinelles Lernen verändert nicht nur, wie wir forschen. Es verändert auch, was wir als gute Erklärung akzeptieren. Ein Modell, das korrekte Vorhersagen trifft, ohne dass wir jeden Schritt nachvollziehen können, fordert unser Verständnis von wissenschaftlicher Erkenntnis heraus. Die künftige Chemie wird nicht weniger experimentell sein. Aber sie wird experimenteller über sich selbst nachdenken müssen: Darüber, welche Daten wir sammeln, wie wir sie kuratieren und wie wir sie mit theoretischem Wissen verbinden.

Darüber hinaus haben LLMs gezeigt, dass Sprache automatisiert werden kann. Auch das beeinflusst die Wissenschaft. Sie eröffnen die Chance, sich besser auf das Wesentliche zu konzentrieren. Allerdings bringen sie die Verantwortung mit sich, das Wesentliche auch als solches zu erkennen und das Modell als das zu begreifen, was es ist: eine statistische Vorhersage erwartbarer Sprache.

Am Ende wird die KI weder das Periodensystem noch das Reagenzglas ersetzen. Aber sie wird verändern, wie wir beides verwenden. Und vielleicht – so meine Hoffnung – werden meine Kolleg:innen dank maschineller Lernverfahren seltener frustrierende Sackgassen beschreiten müssen, um sie als solche zu erkennen. Stattdessen können wir bewusst riskantere Experimente wählen, weil wir die aussichtslosen Wege bereits am Computer ausgeschlossen haben.

Die chemische Forschung wird also nicht leichter werden, nur spannender ...

Maximilian P.-P. Kovar forscht an der Schnittstelle von theoretischer und physikalischer Chemie sowie chemischer künstlicher Intelligenz. Im Rahmen seines DOC-Fellowships der ÖAW untersucht er als Doktorand an der TU Wien die Vorhersagbarkeit chemischer Reaktionen anhand möglichst weniger Stellvertreterreaktionen, die er mittels quantenchemischer Workflows errechnet, um bessere Machine-Learning-Modelle zur Vorhersage chemischer Reaktionen entwickeln zu können. Er studierte Chemie an der Universität Wien sowie Chemie und Technologie der Materialien an der TU Wien und absolvierte einen Forschungsaufenthalt an der Universidade de Santiago de Compostela. In seinen Abschlussarbeiten befasste er sich mit reaktionsbedingtem Ladungstransport, Reaktionsdatensätzen und Protonierungsprozessen sowie in porösen Materialien eingeschlossenen Flüssigkeiten. All diese Themen verbindet der Blick auf chemische Transformationen in Bereichen, die meist statische Molekül- und Materialeigenschaften betrachten. Darüber hinaus engagiert er sich in der Lehre, in universitären Gremien und besonders für den interdisziplinären Austausch zwischen Chemie, Physik, Informatik, Data Science und Mathematik. Maximilian ist seit 2025 PRO SCIENTIA Stipendiat.

Literatur

- Back, Seoin, Alán Aspuru-Guzik, Michele Ceriotti, et al. 2024. "Accelerated Chemical Science with AI." *Digital Discovery* 3 (1): 23–33. <https://doi.org/10.1039/D3DD00213F>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March 3, 610–23. <https://doi.org/10.1145/3442188.3445922>.
- Bengio, Yoshua, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. "Curriculum Learning." *Proceedings of the 26th Annual International Conference on Machine Learning*, June 14, 41–48. <https://doi.org/10.1145/1553374.1553380>.
- De Landsheere, Jasper, Maximilian P. P. Kovar, Konstantin Mark, et al. 2026. "Toward On-the-Fly Prediction of Reaction Energetics for High-Throughput Screening." Preprint, January 9. <https://doi.org/10.26434/chemrxiv-2026-np10c>.
- Heid, Esther, Charles J. McGill, Florence H. Vermeire, and William H. Green. 2023. "Characterizing Uncertainty in Machine Learning for Chemistry." *Journal of Chemical Information and Modeling* 63 (13): 4012–29. <https://doi.org/10.1021/acs.jcim.3c00373>.
- Jumper, John, Richard Evans, Alexander Pritzel, et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* 596 (7873): 583–89. <https://doi.org/10.1038/s41586-021-03819-2>.
- Meuwly, Markus. 2021. "Machine Learning for Chemical Reactions." *Chemical Reviews* 121 (16): 10218–39. <https://doi.org/10.1021/acs.chemrev.1c00033>.
- Smith, Justin S., Ben Nebgen, Nicholas Lubbers, Olexandr Isayev, and Adrian E. Roitberg. 2018. "Less Is More: Sampling Chemical Space with Active Learning." *The Journal of Chemical Physics* 148 (24): 241733. <https://doi.org/10.1063/1.5023802>.
- Strieth-Kalthoff, Felix, Sara Szymkuć, Karol Molga, Alán Aspuru-Guzik, Frank Glorius, and Bartosz A. Grzybowski. 2024. "Artificial Intelligence for Retrosynthetic Planning Needs Both Data and Expert Knowledge." *Journal of the American Chemical Society*, ahead of print, April 10. <https://doi.org/10.1021/jacs.4c00338>.
- Szymkuć, Sara, Ewa P. Gajewska, Tomasz Klucznik, et al. 2016. "Computer-Assisted Synthetic Planning: The End of the Beginning." *Angewandte Chemie International Edition* 55 (20): 5904–37. <https://doi.org/10.1002/anie.201506101>.



Impressum

Studienstiftung PRO SCIENTIA
Lisa Simmel, Geschäftsführerin

Otto Mauer Zentrum
Währinger Str. 2-4/22, 1090 Wien
office@proscientia.at

ISSN: 3061-094X (Print), 3061-0958 (Online)
<https://www.proscientia.at/publikationen/>

Die Verantwortung für den Inhalt der Beiträge
liegt bei der Verfasserin/dem Verfasser.

Radaktion: Christian Jostmann
Satz: Leonie Licht

Cover: Francesca-Maria Raffler, Wien
Druck: Facultas, Stolberggasse 26, 1050 Wien

Zum Cover

Ein Schmetterling als Sinnbild für Metamorphose und Transformation: Auf seinen Flügeln trägt er symbolische Ausschnitte der vielfältigen wissenschaftlichen und künstlerischen Disziplinen, die PRO SCIENTIA ausmachen und die sich in einem stetigen Wandel befinden. Die mechanische Gestaltung des Schmetterlings greift diese Dynamik auf und verweist auf die gegenwärtigen Transformationsprozesse in Wissenschaft, Kunst und Gesellschaft. Dem Entwurf liegt eine handgemachte Collage mit Bleistiftzeichnung zugrunde, als Zeichen menschlichen, künstlerischen Schaffens im Kontrast zur immer mehr dominierenden Verwendung von "KI".

Francesca-Maria Raffler, Wien