

## Pressemitteilung

Gesellschaft Deutscher Chemiker e.V.

Maren Mielck

10.06.2022

<http://idw-online.de/de/news795323>

Forschungsergebnisse, Wissenschaftliche Publikationen  
Chemie, Werkstoffwissenschaften  
überregional



GESELLSCHAFT DEUTSCHER CHEMIKER

## Mehr Daten in der Chemie

### Berichte über negative Ergebnisse würden die Planung von chemischen Synthesen verbessern

Unzählige chemische Experimente sind in Datenbanken zugänglich. Dennoch sind diese Daten nicht gut genug, um mithilfe von künstlicher Intelligenz (KI) und maschinellem Lernen bei neuen Synthesen Produktausbeuten vorherzusagen, hat ein Forschungsteam herausgefunden. Wie das Team in der Zeitschrift Angewandte Chemie berichtet, liegt die schlechte Datenqualität vor allem auch an der Neigung der Wissenschaftler:innen, fehlgeschlagene Experimente nicht zu veröffentlichen.

Bei der Vorhersage von Molekülstrukturen und Materialeigenschaften leisten maschinelle Lernwerkzeuge bereits sehr viel. Geht es aber konkret um die Produktausbeute bei einer Synthese, liefern KI-basierte Modelle nur ungenaue Vorhersagen, stellten Forschende um Frank Glorius von der Westfälischen Wilhelms-Universität Münster fest.

Den Grund für das Scheitern identifizierten die Forschenden in der Qualität der Daten, auf die die KI zurückgreift. „Die Vorhersage von Reaktionsausbeuten, also der Reaktivität, ist viel herausfordernder als die Vorhersage von molekularen Eigenschaften. Ausgangsstoffe, Hilfsstoffe, Mengen, Bedingungen, Ausführung des Experiments – sie alle bestimmen die Ausbeute, was bedeutet, dass dieses Problem sehr datenintensiv wird,“ erklärt Frank Glorius, Korrespondenzautor der Studie. Trotz der riesigen Menge an verfügbarer Literatur reichten die Daten also nicht aus, um Ausbeuten korrekt vorherzusagen.

Das Problem liege dabei häufig nicht an zu wenig berichteten Experimenten, sondern an einer Datenschieflage. Drei mögliche Ursachen identifizierten die Forschenden: Zum einen unterliegen die Ergebnisse von chemischen Synthesen grundsätzlich experimentellen Fehlern. Zweitens treffen Wissenschaftler:innen bei ihrer Reaktionsplanung mehr oder weniger bewusst eine Vorauswahl, die auf persönlicher Erfahrung und Vertrauen in etablierte Verfahren beruht. Und drittens melden sie fehlgeschlagene Reaktionen weniger häufig als erfolgreiche – vermeintlich führen nur Reaktionen mit einem positiven Ergebnis zu einem wissenschaftlichen Fortschritt.

Um herauszufinden, welcher der drei Faktoren den größten Einfluss hat, veränderten Glorius und sein Team den Datensatz für vier gängige (und daher datenreiche) organische Reaktionen. So erhöhten sie künstlich den experimentellen Fehler, reduzierten die Daten auf eine noch kleinere Vorauswahl, oder sie nahmen aus dem Datensatz negative Ergebnisse heraus. Am wenigsten Einfluss auf das Modell hatte der experimentelle Fehler, sehr groß war dagegen der Einfluss der negativen Ergebnisse, heißt es in der Studie.

Wissenschaftler:innen sollten also unbedingt ermutigt werden, fehlgeschlagene Ergebnisse ebenfalls zu berichten, schreibt das Team. Damit verbessere sich die Datenlage, und KI-basierte maschinelle Lernmodelle können die Arbeit beschleunigen und effizienter machen. „Maschinelles Lernen wird die Effizienz in der (molekularen) Chemie dramatisch verbessern. Chemiker:innen kommen mit weniger Reaktionen aus, um ein bestimmtes Ziel zu erreichen wie zum Beispiel optimale Ausbeuten. Dadurch werden chemische Prozesse – und die Welt – nachhaltiger,“ bekräftigt Glorius.

Angewandte Chemie: Presseinfo 12/2022

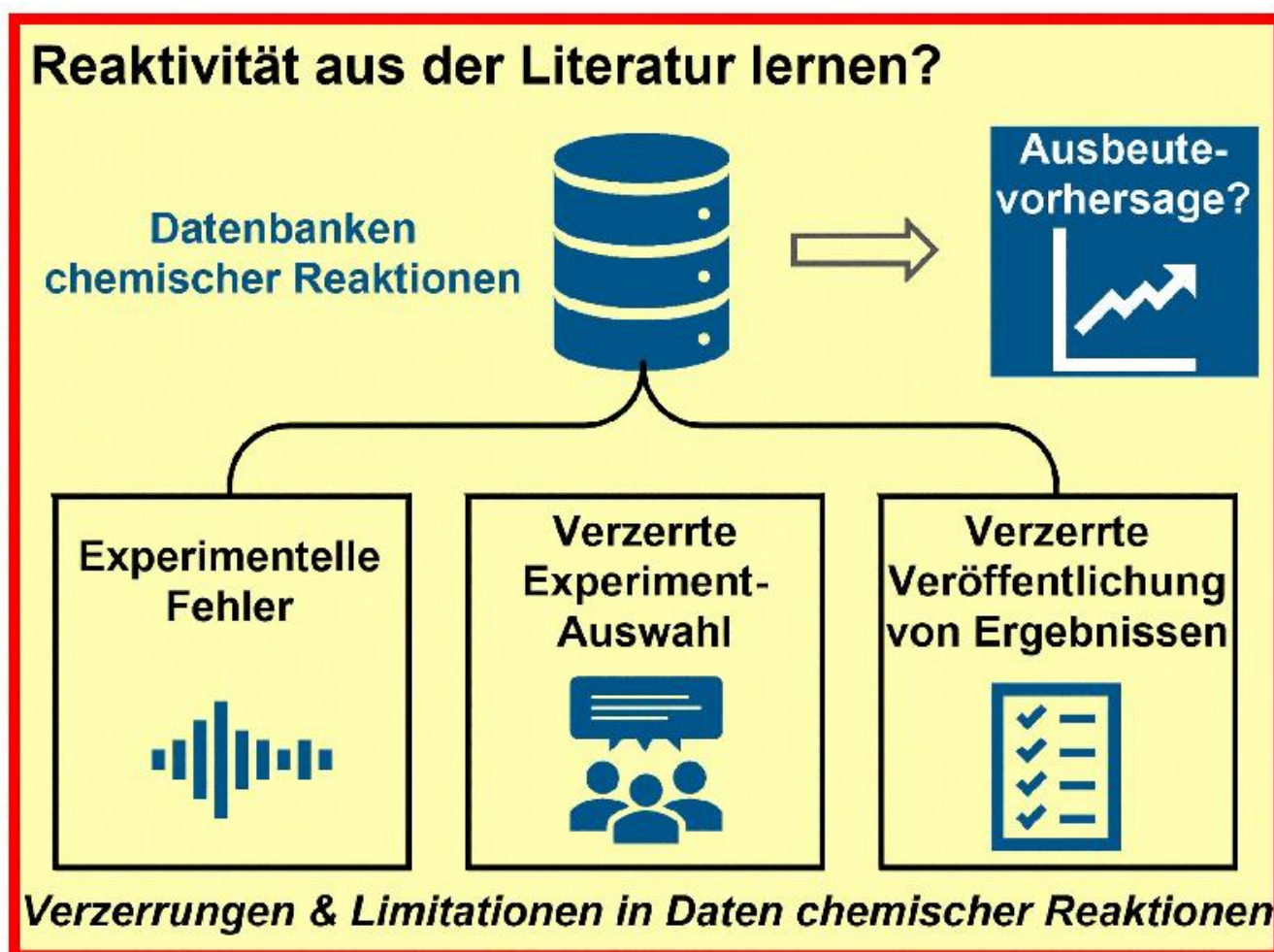
Autor/-in: Frank Glorius, Westfälische Wilhelms-Universität Münster (Germany),  
<https://www.uni-muenster.de/Chemie.oc/glorius/glorius.html>

Angewandte Chemie, Postfach 101161, 69451 Weinheim, Germany.  
Die "Angewandte Chemie" ist eine Publikation der GDCh.

Originalpublikation:

<https://doi.org/10.1002/ange.202204647>

URL zur Pressemitteilung: <http://presse.angewandte.de>



Mehr Daten in der Chemie  
(c) Wiley-VCH

D