

Biologie und Künstliche Intelligenz als natürliche Partner

Kein Tag vergeht ohne Berichte, Kommentare und Beiträge zu den Chancen und Risiken der Anwendung künstlicher Intelligenz (KI) in allen Lebensbereichen. Inwieweit betrifft dies die Biologie? Die Lebenswissenschaften und diverse Ansätze generativen maschinellen Lernens sind ideale Partner, denn ihre zielgerichtete Kombination eröffnet fundamental neue Einsichten und verspricht eminenten Erkenntnisgewinn. Aber auch im engen und weiten Umfeld der Wissenschaft ändert KI die Praktiken. Diese Entwicklung ermöglicht neue Herangehensweisen beim Planen von Experimenten, Extrahieren von Methoden, Erheben von aggregierten Daten oder beim Publizieren und betrifft die gesamte Wissenschaft und Wissenswelt, aber insbesondere die Biologie.

Unter den Naturwissenschaften vertritt die Biologie eine einzigartige Breite an Komplexitätsebenen – möglicherweise nur übertroffen von der Astronomie. An dem einen Ende der Skala befassen sich die Biochemie und Biophysik nanoskopisch mit einzelnen molekularen Prozessen. Am anderen Ende beschreibt die Ökosystembiologie die großräumigen Wechselbeziehungen innerhalb der Biosphäre. Je nach Betrachtung können zahlreiche Ebenen unterschieden werden, die auf jeder höheren Ebene neue Eigenschaften aufweisen. Das Charakteristikum jeder Ebene sind die unermessliche Vielzahl und Unterschiedlichkeit der zu berücksichtigenden Wechselwirkungen von Molekülen, Zellen, Organen, Organismen, Populationen und Ökosystemen.

Molekulare Wechselwirkungen und KI in der Zellbiologie

Wenn wir eines Tages alle relevanten molekularen Interaktionen in einer Zelle und den subzellulären Kompartimenten vorhersagen und modellieren können, dann hätten wir die Zellbiologie tief durchdrungen, ja möglicherweise verstanden. Betrachten wir eine typische kleine eukaryotische Zelle der Dimension von $10 \times 4 \times 5 \mu\text{m}^3$. Diese Zelle mag 600 Mio. Polypeptidmoleküle, 2 Mio. Nukleinsäuremoleküle (deren Größe von Zigmillionen zu einigen wenigen Nukleotiden reicht), 30 Mrd. Metabolite und Ionen sowie 5 Billionen Wassermoleküle enthalten. Die Zahl der unterschiedlichen Polypeptide nimmt stark zu, wenn alternative Spleißvarianten und posttranslationale Modifikationen berücksichtigt werden. Insoweit werden Analysen möglicher molekularer Interaktionen und kausaler Zusammenhänge bislang meist extrem reduktionistisch angelegt.

Die Durchdringung dieser Wechselwirkungen verlangt nach neuen Ansätzen und wird durch maschinelles Lernen wesentlich profitieren. Mit der Proteinmodellierung durch AlphaFold [1] hat es einen ersten Quantensprung an überzeugenden Strukturvorhersagen für Proteine durch KI gegeben, die nun um die Komponente der wechselseitigen Bindung – beispielsweise eines Wirkstoffes an ein Zielprotein – erweitert wurde [2]. Insoweit überrascht es nicht, dass diese Fortschritte als beeindruckendes Beispiel erfolgreicher KI-Anwendung in der Biologie herangezogen werden. Die Weiterentwicklung solcher KI-Anwendungen in Kombination mit zellulä-



Prof. Dr. Karl-Josef Dietz von der Universität Bielefeld ist seit 2020 Präsident des VBIO.

ren Modellen wird uns Schritt für Schritt zu einem umfassenderen Verständnis der Zellbiologie führen.

Wie weit sind wir wirklich?

Eine Recherche in der Datenbank *Web of Science* mit den Suchbegriffen *Artificial Intelligence* und *Bio** findet mehr als 20.000 Publikationen, von denen sich viele mit medizinischer Diagnostik befassen, aber auch mit Biodiversität, Mikroskopie, Tieren, Verhaltensbiologie oder zellulärer Signalverarbeitung. Allerdings stellen viele der Artikel Meinungsbilder und Prognosen vor und nur ein kleinerer Anteil erfolgreiche KI-Projekte.

Erfolgreiche Anwendungen der KI in der Biologie betreffen vor allem die Objekt- und Mustererkennung und die Verarbeitung von „Big Data“ und sind letztendlich konsequente Weiterentwicklungen der Bioinformatik mit zusätzlichen (automatisierten) Rückkopplungsmechanismen, die zur Weiterentwicklung der Programme führen. Trainierte Objekterkennung in mikroskopischen oder makroskopischen Bildern, Erkennen von Verhaltensregeln anhand von beispielsweise satellitengestütztem Individuen-Tracking im Freiland, Modellierung molekularer oder individueller Prozesse sind Anwendungsfelder, deren Vielfältigkeit beindruckend ist.

Interaktion von Organismen als Feld für KI-Forschung

Ein weiteres Beispiel für zukünftige KI-Anwendungen sind die diversen organismischen Wechselwirkungen in der Natur, sei es in der Rhizosphäre oder der Darmflora. Ein Anteil von 15 Prozent der Weltbiomasse wird den Bakterien zugeschrieben, zwei Prozent den Pilzen, ein Prozent den Protozoen und nur 0,3 Prozent den Tieren [3]. Den Rest stellen die Pflanzen. In einem Gramm Bodenprobe befinden sich vielfach mehr als 50.000 Arten. Die Identifizierung der Spezies und Abschätzungen ihrer Menge gelingen zunehmend besser mit metagenomischen Ansätzen, die die Gesamtheit der Erbinformation in einem Lebensraum erfassen. Nächstes Ziel wird sein, die positiven, negativen und neutralen Wechselbeziehungen zwischen den Organismen zu beschreiben.

Wie im Beispiel der molekularen Wechselwirkungen untersuchen wir derzeit stark vereinfachte Systeme. Ein umfassenderes Verständnis der vielfältigen Interaktionen zwischen den Arten im Boden inklusive der Pflanzenwurzeln werden KI-Ansätze eröffnen. Diese Erkenntnisse werden die Entwicklung von Strategien unterstützen, um etwa die Pflanzengesundheit in der Landwirtschaft zu erhöhen. KI-basierte Ansätze werden uns in die Lage versetzen, die eingangs dargestellten Komplexitätsebenen besser miteinander zu verknüpfen – beispielsweise die molekularen Zellmodelle mit den organismischen Interaktionen.

Prognose zur KI im wissenschaftlichen Erkenntnisgewinn

Ein Editorial erlaubt es, Meinungen zu formulieren und den Anstoß zur Diskussion zu geben. Abbildung 1 zeigt den Ablauf eines For-

schungsvorhabens von der anfänglichen Bestandsaufnahme bis hin zur Veröffentlichung neuer Erkenntnisse. Die sechs Kästchen rechts neben den Arbeitsschritten zeigen meine Einschätzung zu der Frage, welchen Stellenwert KI-basierte Verfahren kurzfristig (rot) und mittelfristig (grün) einnehmen könnten. Graue Kästchen geben an, dass meines Erachtens KI qualitativ und quantitativ auf absehbare Zeit keine Rolle in der Spitzenforschung spielen wird. Bereits beim Entstehen dieses Editorials diskutierten Kolleg/-innen und ich über die naheliegendste Verteilung der roten, grünen und grauen Kästchen – diskutieren Sie doch bitte mit (→ karl.josef.dietz@uni-bielefeld.de).

Kurz- und mittelfristiger Einsatz von KI-basierten Methoden

Trainierbare KI-basierte Programme zur Bildanalyse und Mustererkennung werden, wie oben ausgeführt, kurzfristig zum allgemein genutzten Mittel der Wahl der Biowissenschaften. Die Bestandsaufnahme des Wissens in der anfänglichen und begleitenden Recherche, die Methodenwahl für die durchzuführenden Experimente, der Vergleich mit verfügbaren (großen) Datensätzen und die Abfassung der Veröffentlichungstexte werden ebenfalls kurzfristig von KI profitieren. Bereits heute extrahieren Datenbanken experimentelle Methoden aus Publikationen und stellen diesen Fundus teils gegen Entgelt zur Verfügung. Die Autor/-innen werden um Prüfung gebeten, was entweder zeigt, dass publizierte Methoden leider häufig unvollständig oder die extrahierten Methodenbeschreibungen fehlerhaft sein können. Pubcompare.ai wirbt mit mehr als 40 Millionen experimentellen Protokollen [4]. Verlage entwickeln Verfahren, um die Nutzung großer Sprachprogramme (*Large Language Models*, LLM wie ChatGPT4) in eingereichten Publikationen nachzuweisen. Die Verlage sind zuversichtlich, dies zu erreichen, wobei sich die Frage stellt, ob verbesserte KI-basierte Programme nicht der Aufdeckung durch die aktuellen Analyseprogrammen entkommen.

In wesentlichen Prozessen der Wissenschaft (Abbildung 1) kann mittelfristig eine zunehmende Bedeutung von KI-Ansätzen prognostiziert werden. Die grünen Kästchen beschreiben diese Perspektive. Es wird in den kommenden 5–10 Jahren spannend werden, die tatsächliche Entwicklung zu beobachten.

Aber wie sieht es mit dem zentralen Element biowissenschaftlicher Forschung, der Hypothesenbildung, aus? Drei der sechs Kästchen habe ich grün eingefärbt. Bereits heute gibt es viele Forschungsprojekte, deren Ausgangshypothesen eher oberflächlich sind. Insofern wird KI dazu beitragen können, solche Hypothesen zu verfeinern. Weniger wahrscheinlich ist dies bei innovativen und originellen Hypothesen.

Regeln für KI-basierte Ansätze

Ein Charakteristikum generativen maschinellen Lernens ist ein breiter Lösungsraum, so dass bei wiederholter Eingabe derselben Daten

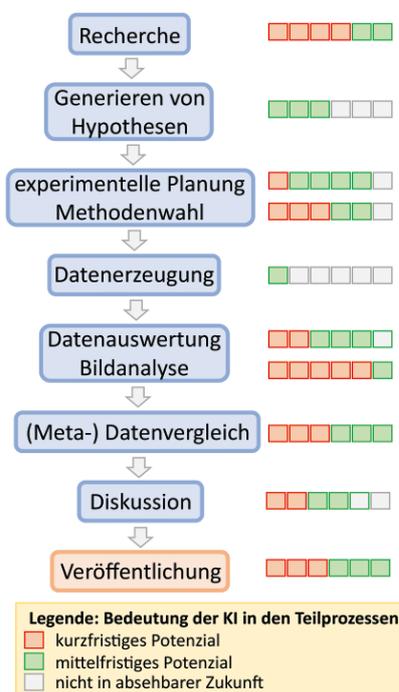


Abb. 1 Mögliche Bedeutung von KI in den einzelnen Schritten des wissenschaftlichen Erkenntnisprozesses

in das Programm unterschiedliche Ergebnisse erzielt werden. Der sorgfältige Umgang mit dieser Variabilität wird neue Regeln guten wissenschaftlichen Arbeitens verlangen. Der Generalsekretär der Vereinten Nationen António Guterres fasste dies beim *AI for Good Global Summit* am 6. Juli 2023 [5] mit drei prinzipiellen Forderungen zusammen: Einhaltung der Menschenrechte, volle Transparenz und Rechenschaftspflicht. Dies gilt für die Anwendungen der KI in der Wissenschaft in gleicher Weise.

Interessant ist, dass KI dazu beitragen könnte, die Wissenschaft zu demokratisieren, da die Entwicklung generativer KI auf verfügbaren Datensätzen aufbauen kann und somit entkoppelt von der ressourcenintensiven Durchführung von Experimenten oder Analysen ist. Hier fügen sich *Open Data*, *Open Science* und KI zu einer vielversprechenden Perspektive zusammen. Dazu sollten möglichst bald die Curricula in der Oberstufe und in den betroffenen Studienfächern angepasst werden, um mit den Chancen, Anwendungen und Risiken von KI vertraut zu machen.

Die Risiken, die mit der Anwendung von KI in der Wissenschaft verbunden sind, sind vielfältig und reichen von ungeprüfter Datenübernahme, fehlender Offenlegung der Algorithmen, Programmstrukturen und Trainingsdaten bis hin zur Enteignung der Forschenden durch fehlende Zitierung und Marginalisierung der Einzelbefunde. Und schließlich kann es zu einer KI-Degeneration kommen, wenn KI-generierte Aussagen von KI-Programmen weiterverwendet werden und dadurch an Sinn und Bezug verlieren. Trotzdem gehören KI-Anwendungen zu dem zu begrüßenden Repertoire der Wissenschaft, denn KI wird bei aller Kritik zu einer Beschleunigung von Entwicklung in vielen Bereichen führen, die Vielfalt und die Differenziertheit neuer Erkenntnisse beträchtlich erhöhen und insbesondere zur Lösung komplexer Probleme erheblich beitragen.

Ihr

Anmerkung nach Drucklegung: Die Aktualität dieser Thematik wurde durch die Vergabe des Physik-Nobelpreises an Geoffrey Hinton und John Hopfield am 8.10.2024 für ihre KI-Grundlagenforschung sowie des Chemie-Nobelpreises an die Proteinforscher David Baker, Demis Hassabis und John Jumper am 9.10.2024 unterstrichen.

Literatur

- [1] J. M. Jumper et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589.
- [2] J. Abramson et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500.
- [3] Y. M. Bar-On et al. (2018). The biomass distribution on Earth. *Proc. Natl Acad. Sci. USA* 115, 6506–6511.
- [4] Pubcompare.ai, <https://www.pubcompare.ai/terms-of-services/>
- [5] A. G. Hessami et al. (2024). Artificial Intelligence for the Benefit of Everyone. *Computer* 57 (9), 68–79, <https://doi.org/10.1109/MC.2024.3411797>