

Das Zusammenspiel von Mensch und Maschine erreicht mit den künstlichen Intelligenzen neue Dimensionen. In der Wissenschaft suchen Algorithmen bereits neue Theorien.

Wenn Kunst mit KI mit Kunst mit KI

Das ist doch ein Selbstporträt von Frida Kahlo. Wenn auch eine seltsame Version. Der Walliser Künstler Jonas Wyssen hat einige Meisterwerke der Kunstgeschichte den KI-Algorithmen ausgesetzt respektive seinem ausgeklügelten Eingabecode für KI-Algorithmen. Wyssens Ziel: die Bilder und Stile bewusst zu kopieren und zu imitieren – in der Kunst nennt sich dieser Vorgang Appropriation –, dabei aber stets einen «Fehler in der Matrix» einzubauen. Die surrealen, traumgleichen Kreationen verwirren: Was ist echt? Was kommt von der KI? Was vom Künstler?

Bild: Jonas Wyssen



Maschine, schreib mir das schnell!

Ob Publikationen oder studentische Arbeiten, in der Wissenschaft gibt es zunehmend Texte, die mit Hilfe von künstlicher Intelligenz entstanden sind. Meinungen zum Umgang damit gehen auseinander.

Text Michael Baumann

Das kluge Textprogramm Chat-GPT hat seit einigen Monaten Hochkonjunktur, unter anderem weil es von der Firma Open AI kostenlos zugänglich gemacht wurde. Auch im akademischen Bereich eröffnen sich damit ganz neue Möglichkeiten, wobei die Resultate erstaunlich gut sind: beim Recherchieren, Schreiben und Reviewen – oder auch beim Betrügen. Die Wissenschaftszeitschrift Nature jedenfalls hat bereits Regeln für den Umgang mit KI erlassen. Unter anderem muss von den Autoren und Autorinnen die Herkunft einer maschinell entstandenen Aussage transparent dargelegt werden.

Der Wissenschaftshistoriker Mathias Grote von der Universität Bielefeld diagnostiziert einen Zeitpunkt des Wandels: «Es scheint einer jener Momente erreicht, die vor allem in der Nutzung einen Durchbruch darstellen.» Vergleichbare Zeitpunkte habe es beim Internet, bei der Suchmaschine von Google oder beim Smartphone gegeben. Dass sich KI dereinst durchsetzen werde, habe sich seit geraumer Zeit angekündigt. Grote erinnert sich an Debatten Anfang der 2000er-Jahre: «Damals ging es noch mehr um eine kategorische Klärung und um eine theoretische Diskussion.» Darum also, was künstliche Intelligenz eigentlich ist. Heute dagegen steht die Technologie selbst im Vordergrund. Und natürlich spielen ökonomische und politische Motive bei der Entwicklung eine wichtige Rolle: Datenkonzerne und Grossverlage etwa würden mit den Textprogrammen monetäre Interessen verfolgen.

Ende des selbstständigen Denkens?

Was seit Chat-GPT-4 für alle zugänglich ist, wird auch beim Verfassen von wissenschaftlichen Texten verwendet. Entsetzt und erschüttert über die neuen Werkzeuge, die «Internet-Unternehmen den Studierenden und Forschenden zur Verfügung stellen», zeigt sich Kardiologe Matthias Barton von der Universität Zürich. Er ist Mitherausgeber der Open-Access-Fachzeitschrift für Biomedizin und Biowissenschaften E-Life. «Es ist nicht ganz klar, wer hinter diesen Programmen steckt», erklärt er. Zudem würden diese mit sehr fragwürdigen Aussagen beworben wie zum Beispiel, dass künstliche Intelligenz es einem Computer ermögliche, ohne menschliches Zutun wie ein Mensch zu denken. KI sei überhaupt ein falscher Begriff, findet er, spricht sogar von einem Etikettenschwindel. Von Intelligenz könne keine Rede sein, es gehe ganz banal um den

Einsatz von Algorithmen, insbesondere im Publikationssystem. Grundsätzlich finde dort eine Entpersonalisierung statt, sagt Barton. Und das selbstständige Denken rücke in den Hintergrund. «Dieser Umstand führt dazu, dass Studierende und junge Wissenschaftler, die wissenschaftliche Arbeiten oder Teile davon mit KI erstellen können, jedoch ihren eigenen Namen als Autor angeben.»

Nicht zu viel Kontrolle

Um dieser neuen Form des Betrügens bereits im Studium beizukommen, braucht es auch neue Herangehensweisen. Dozierende müssten den Studierenden das Ethos der Selbstkontrolle vermitteln, schlägt der St. Galler Wissenschaftshistoriker Caspar Hirschi vor. Für eine eigentliche Überprüfung von studentischen Arbeiten brauche es nämlich starke Verdachtsmomente wie etwa grosse Unterschiede bei der Sprachfertigkeit innerhalb einer Arbeit. Mehr Überprüfung wäre allein wegen der schieren Anzahl zu kontrollierender Texte kaum machbar. «Gegenseitiges Vertrauen ist wichtig», ist Hirschi deswegen überzeugt, «verbunden mit dem Einfordern von Selbstdeklarationen, wie das Nature macht.»

Gegen das Betrügen in Fachzeitschriften reicht das wohl nicht. Matthias Barton plädiert für Richtlinien. «Aktiv werden müssten die Schweizerischen Akademien der Wissenschaften, besonders in der Medizin – und zwar schnell.» Er schlägt vor, dass alle Autorinnen schriftlich bestätigen müssen, dass sie keine KI verwendet haben. Für Hirschi ist das zwar auch vorstellbar, er sieht darin aber keinen Gamechanger. «Das System der modernen Wissenschaft basiert auf einem geschulten Vertrauen in die seriöse Arbeit von anderen», erklärt er. In der experimentellen Forschung sei das immer so gewesen. Die Wissenschaft habe ihre Kontrollmechanismen nie flächendeckend anwenden können. Sonst wäre das System kollabiert.

Ähnliche Probleme kennt die Forschung ja eigentlich schon länger, nämlich von Arbeiten, die von Ghostwriterinnen geschrieben wurden. Und vom Kampf gegen Plagiate. In Zeiten von KI-Programmen sei es besonders wichtig, eine originäre Denkleistung einzufordern, ist Wissenschaftshistoriker Mathias Grote dennoch überzeugt. Allenfalls wäre ein digitales Wasserzeichen ein Teil der Problemlösung, also eine technische Markierung in Texten zur

Nachverfolgung von Quellen. Um Fälschungen, die mit der Unterstützung von KI entstanden sind, zu erkennen, müsse aber auch schlicht der wissenschaftliche Umgang mit ihr geübt werden, was Teil der studentischen Ausbildung werden müsse. Verbote erachtet Grote dagegen als kontraproduktiv.

Gebrauchstexte im Wikipedia-Style

Die Anwendung von KI in wissenschaftlichen Publikationen wirft aber auch Fragen jenseits ihrer Deklaration auf. Ein Abstract zu verfassen etwa – und hier werden kluge Textprogramme schon häufig eingesetzt – ist laut Matthias Barton eigentlich Teil eines Prozesses, um wissenschaftliches Schreiben zu lernen. Diese Lernerfahrung geht mit dem Einsatz von KI verloren. Grundsätzlich leiste eine Publikation einen kleinen Beitrag zum Stand des Wissens, um die Wissenschaft als Ganzes weiterzubringen. «Aber es besteht ein grosser Druck zum Publizieren, der von den Universitäten ausgeht, dieser kommt den Verlagen äusserst gelegen.» Und er habe zur Folge, dass auch Doktorierende und Postdocs dazu verleitet werden können, mit KI-Programmen zu arbeiten, statt selbstständig Arbeiten anderer Wissenschaftler zu lesen und so auch das wissenschaftliche Schreiben zu lernen.

«Leider geschieht dies häufig in der Medizin», sagt Barton. «Ich weiss von maschinellen Übersetzungen durch die KI DeepL, die eins zu eins in Manuskripte hineinkopiert wurden, was natürlich nicht angegeben wurde.» Auch hier fehle es an nötigen Regulierungen, wie sie kürzlich der deutsche Ethikrat oder verschiedene britische Universitäten erlassen haben. «Denn der Mensch ist von Natur aus bequem und nutzt, was ihm zur Verfügung steht.»

Caspar Hirschi findet: «Man kann mit KI einen Text schreiben lassen, der gar nicht so schlecht ist.» Die Standardargumente bringe sie zusammen, vergleichbar mit Wikipedia. Aber ein solcher Text bilde nur einen Wissensstand ab, sei nicht immer referenziert und bar jeder Originalität. Darüber hinaus seien eine Forschungsleistung und die Herstellung von Zusammenhängen nicht möglich. Neues methodisches Vorgehen sei von der Forschung zu leisten und könne nicht an eine KI delegiert werden. «Dadurch wird die Kluft zwischen neuer Forschung und einfachem Zusammenfassen grösser.»

Mathias Grote beurteilt das Vermögen von KI ähnlich. Mit ihr seien Gebrauchstexte zwar schnell und effektiv zu erstellen und Recherchen durchzuführen. «Das meiste stimmt, wenn sich das Programm auf hinlänglich bekannte und verfügbare Fakten abstützen kann, aber eben doch nicht alles.» Deshalb müssten solche Texte zwingend weiterbearbeitet, überarbeitet und korrigiert werden und könnten höchstens als Grundlage dienen. «Ich kann mir schwerlich vorstellen, dass genuine wissenschaftliche Entdeckungen jemals nur von einer KI gemacht werden.» KI

sei vielmehr als sinnvolles Hilfsmittel einzusetzen, um Standardinformationen schnell abrufen und Routinen sowie Berechnungen durchführen zu können – ähnlich wie es früher Handbuch, Tabellen und Taschenrechner waren.

Bei aller Kritik an der Verwendung von KI-Programmen ortet auch Matthias Barton hier Chancen. «Wenn der Algorithmus gut programmiert ist, lassen sich Recherchen an die KI delegieren, was Zeit und Aufwand spart», führt er aus. Doch die gelieferten Ergebnisse müssten zwingend geprüft werden, sonst überwiege die Gefahr. In der Medizin könne KI in beschränktem Mass bei der Erfassung und Verarbeitung von grossen Datenmengen von Vorteil sein. Wenn aber in Forschungsanträgen die Verwendung von KI explizit aufgeführt werde, nur weil sie gerade en vogue sei, dann stimme etwas nicht. «Besser wäre es, einmal inezuhalten und zu überlegen, ob von einem Fortschritt gesprochen werden kann.»

Auch im Peer-Review-Prozess von Fachzeitschriften wird KI bereits verwendet. Der Mitherausgeber von E-Life, Matthias Barton, sieht sie als nicht geeignet dafür an, da

KI nicht über Fachwissen verfüge. Auch könne sie die inhaltliche Relevanz der in wissenschaftlichen Arbeiten zitierten Literaturstellen gar nicht bewerten. Bei E-Life kämen grundsätzlich nur ausgewiesene Fachleute als Gutachtende in Betracht. Trotzdem seien solche Programme beim Peer-Review als Hilfsmittel bereits weit verbreitet. Sie werden von vielen Verlagen verwendet, um etwa Editoren

mögliche Gutachter vorzuschlagen. «Sie sind froh, wenn sie überhaupt welche finden.» Doch auch sonst im Peer-Review-Prozess wird auf KI gesetzt. Denn selbst wenn eine Gutachterin gefunden ist, ist sie in der Regel überlastet und muss sich die Zeit zwischen regulärer Arbeit und Begutachtung aufteilen. KI kann zum Beispiel diesen Stress mildern, indem sie ihr auf der Grundlage ihrer früheren Arbeiten, Kommentare und Interaktionen Vorschläge unterbreitet. Das zumindest versprechen Anbieter solcher Programme.

Dass Unternehmen wie der Elsevier-Verlag oder Online-Publisher wie MDPI im Peer-Review auf KI setzen, überrascht Barton nicht. «Der Publishing-Prozess ist von vielen Firmen längst in Länder wie China, Indien oder Pakistan ausgelagert worden.» Durch äusserst billige Arbeitskräfte sparten die Verlage enorme Kosten, reduzierten aber gleichzeitig auch die Verantwortung für den Inhalt der von ihnen veröffentlichten Zeitschriften. Je offener das System sei, desto anfälliger sei es. Jeder Algorithmus habe naturgemäss seine Grenzen und Ungenauigkeiten, die dann in den wissenschaftlichen Prozess einflössen.

Michael Baumann ist freier Journalist in Zürich.

«Die Kluft zwischen neuer Forschung und einfachem Zusammenfassen wird grösser.»

Caspar Hirschi



Eine Rakete durchbricht die stille Idylle von Claude Monets erstem impressionistischem Gemälde «Impression, Soleil Levant» (1872). Man ahnt den Untergang von Natur im Leuchten des Sonnenaufgangs.

Bild: Jonas Wyssen



«Wir sollten Mensch und KI nicht gegeneinander ausspielen»

Herauszufinden, wie gewisse selbstlernende Algorithmen funktionieren, verursacht viel mühevollen Arbeit, meint Claus Beisbart. Berührungspunkte damit kennt der Wissenschaftsphilosoph dennoch keine – im Gegenteil, sie seien bereits jetzt ein Werkzeug in seiner Disziplin.

Text Florian Fisch Foto Raffael Waldner

Haben Sie schon versucht, mit einem Chatbot eine philosophische Diskussion zu führen?

Ja, wir haben diskutiert, wie der Mensch seine Autonomie, seine Eigenständigkeit bewahren kann, wenn künstliche Intelligenz oder KI immer leistungsfähiger wird. Chat-GPT hat durchaus einschlägige Begriffe verwendet wie menschliche Würde, Transparenz und Fairness. Von einem tiefen philosophischen Gespräch sind wir aber noch weit entfernt. Der Chatbot hat sich zum Beispiel kaum positioniert. Aber sonst wenden wir die KI in der Philosophie schon an.

Wirklich? Wie dient Ihnen die KI?

Wir simulieren zum Beispiel, ob eine philosophische Methode zu Konsens führen kann, wenn es um Streitthemen wie Fleischkonsum geht. Das hängt natürlich stark von der Struktur der Debatte ab. Es gibt bei uns jedenfalls keine Ängste, mit dem Computer zu arbeiten.

Kann man KI als Blackbox bezeichnen, die derart komplizierte Berechnungen anstellt, dass am Schluss niemand mehr versteht, was darin eigentlich passiert?

Das ist sicher ein Charakterzug von neuer KI, aber nicht die Definition. Es gibt auch die gute altmodische KI. Dazu gehören beispielsweise Simulationen, wie sie in den Klimawissenschaften angewandt werden. Die Regeln, nach denen diese KI arbeitet, die Gleichungen, werden von den Forschenden vorgegeben. Sie wissen daher, mit welchen Größen der Computer rechnet. Anders ist es bei selbstlernenden Programmen. Diese geben sich die Regeln auf der Basis von Daten selbst. Ihre Anwendungen sind besonders schwierig zu durchschauen, nachzuvollziehen und kritisch zu überprüfen.

Können Sie ein konkretes Beispiel dafür nennen?

Ein neuronales Netz wird erst mal trainiert. Dazu bekommt es Daten, die korrekt klassi-



fiziert sind. Wenn es dann Bilder von Hunden und Katzen unterscheidet, wissen wir nicht, auf welche Eigenschaften im Bild es anspricht. Möglicherweise beachtet es vor allem den Hintergrund. Die Forschung wird zwar immer besser darin, nachzuvollziehen, worauf die neuronalen Netze reagieren, aber es bleibt oft ein Rumprobieren – eine unheimlich mühevollen Arbeit.

Kann ein neuronales Netz etwas verstehen? Zum Beispiel Chat-GPT?

Dies könnte schon irgendwann einmal der Fall sein. Aber hinter Chat-GPT steckt letztlich nur ein Sprachmodell, das darauf basiert, welche Wörter mit hoher Wahrscheinlichkeit aufeinander folgen. Der Bot plappert also einfach nach, was am häufigsten im Internet gesagt wurde. Das könnte zwar ausreichen, um den Turing-Test zu bestehen, bei dem ein Mensch entscheiden muss, ob er mit einem Computer oder einem anderen Menschen spricht. Aber ein Sprechverhalten, das nach aussen akzeptabel erscheint, heisst nicht unbedingt Verstehen. Dazu gibt es ein Gedankenexperiment von John Searle. Da sitzt eine Person, die kein Chinesisch versteht, in einem geschlossenen Raum, erhält chinesische Texte und chinesische Fragen dazu. Sie verfügt über ein Handbuch mit den Regeln für den Umgang mit den Zeichen. Sie kann also Antworten formulieren, ohne die Sprache zu verstehen.

Was heisst überhaupt verstehen – ich meine bei einem Menschen?

Viele denken da erst mal an den Aha-Effekt: Es macht klick, und man denkt, man hat es verstanden. Aber dieses Gefühl ist individuell und kann trügerisch sein. Daher bauen wir in der Philosophie nicht darauf, sondern versuchen, Verstehen an Fähigkeiten festzu-

machen. Bei der Sprache heisst das dann, dass ich die Bedeutung eines Satzes erklären kann, dass ich weiss, wie ich Ausdrücke neu kombinieren kann. Beim Verstehen von Phänomenen wie beispielsweise der Französischen Revolution kommt es darauf an, Informationen zu vernetzen, Verbindungen zu erkennen und eigenständig Schlüsse zu ziehen.

Wie beurteilen Sie die Fähigkeiten von Chat-GPT?

Chat-GPT bildet eine Art Durchschnitt davon, was im Internet steht, und gewichtet Aussagen nicht nach der Glaubwürdigkeit von Quellen. Diese Fähigkeit könnte man natürlich auch noch einbauen. Im Unterschied zu Menschen brauchen Computer auch unglaublich viele Daten. Einem Kind zeige ich ein einziges Mal einen Puppenwagen, und von da an erkennt es diesen.

Können wir uns aber auf die Outputs einer KI verlassen?

Wenn ich mit neuronalen Netzen Galaxien klassifiziere, dann sehe ich, dass dies in der Vergangenheit funktioniert hat, und vertraue dem Track Record. Das ist bei einem neuronalen Netz die einzige Begründung. Bei der guten altmodischen Klimasimulation dagegen weiss ich mehr: Wir verstehen heute die fundamentalen Prozesse in der Atmosphäre so genau, dass wir auch Szenarien berechnen können, die in den letzten 10 000 Jahren nicht vorkamen.

Im Jahr 2009 machten die Laborautomaten Adam und Eve Furore, die selbst Hypothesen aufstellten und Experimente mit Hefezellen durchführten. Heute hört man nicht mehr viel davon. Werden die Forschenden doch nicht so schnell ersetzt?

Adam und Eve demonstrierten erfolgreich, was möglich ist. Aber die Hypothesen, die Adam aufstellte, folgten alle demselben einfachen Schema. Es gibt schon weitaus Komplexeres: Im letzten Jahr hat man eine KI mit Nasa-Daten über die Planetenbewegungen gefüttert, damit sie daraus ein Gesetz für die Gravitationskraft ableitet wie Newton vor 350 Jahren. Das hat gut funktioniert. Allerdings gab man einen gewissen Rahmen vor. Newton musste gleichzeitig auch noch die Differenzialrechnung erfinden. Aber der Trend geht sicher in diese Richtung. Trotzdem: In näherer Zukunft wird es auf jeden Fall genug Forschungsarbeit für Menschen geben. KI verursacht auch viel Arbeit, weil wir sie nicht verstehen.

Wann man von künstlicher Intelligenz spricht

Für Philosophen wie Claus Beisbart geht es bei künstlicher Intelligenz (KI) darum, rationales **Denken und Handeln von Menschen nachzuzahlen oder sogar zu übertreffen**. Wenn das dabei entstehende System tatsächlich intelligent ist, sprechen Philosophen von starker KI. Wenn es intelligentes Verhalten nur simuliert, heisst das schwache KI.

Im Zentrum der heutigen KI-Forschung steht vor allem das sogenannte maschinelle Lernen, bei dem sich die Algorithmen selbst verbessern. Die Regeln, die sie dabei anwenden, sind **schwer durchschaubar**. Dabei haben sich sogenannte künstliche neuronale Netze als besonders erfolgreich erwiesen. Diese sind im Computer programmierte, **dem Hirn nachempfundene Strukturen**, die selbst lernen. Für Netze mit vielen Schichten hat sich der Begriff Tiefes Lernen durchgesetzt. Solche Netze kommen auch bei der sprachlichen Datenverarbeitung von Chatbots wie Chat-GPT-4 zum Einsatz.

Müssen sich Forschende keine Sorgen machen, dass sie arbeitslos werden?

Der Überraschungscharakter von KI ist klar vorhanden. Aber vielleicht sollten wir Mensch und KI nicht gegeneinander ausspielen. Es ist nämlich schon die Frage: Wer ist eigentlich das Subjekt in der Forschung? Der Mensch? Manche sagen, das stimme schon länger nicht mehr, es seien Gruppen von Menschen, die etwas verstehen. Vielleicht entsteht jetzt ein komplexes Erkenntnisobjekt: der Mensch und sein Computer.

Wird es mal den KI-Philosophen geben?

Grundsätzlich muss ich als Philosoph offen sein. Ich kann nicht ausschliessen, dass es mal eine KI gibt, die bessere Papers schreibt als ich. Für gewisse Dinge wie Logik eignet sich der Computer sowieso sehr gut. Es gibt aber auch philosophische Methoden, für die ich mir den Computer nur schwer vorstellen kann, etwa die Analyse der eigenen Erfahrung. Ausserdem geht es nicht nur darum, schöne Papers zu schreiben, sondern auch, an den relevanten Themen der Zukunft dran zu sein. Da dürfte der Mensch im Vorteil sein. Insgesamt sehe ich es positiv: Mit dem Computer zusammen können wir in der Forschung Dinge machen, die bisher nicht möglich waren.

Florian Fisch ist Co-Redaktionsleiter bei Horizonte.

Von Formeln und Ideen, die die Welt ausmachen

Claus Beisbart (52) publiziert sowohl zu Wissenschaft und Öffentlichkeit als auch zum **Verständnis physikalischer Theorien**. Er studierte Mathematik, Physik und Philosophie und promovierte an der Ludwig-Maximilians-Universität München in Kosmologie und später in Philosophie. Heute ist er **Professor für Wissenschaftsphilosophie** an der Universität Bern und befasst sich unter anderem mit Tiefem Lernen.



Ein DJ bringt Unruhe in das kubistische Werk der russischen Malerin Ljubow Popowa aus dem frühen 20. Jahrhundert.

Bild: Jonas Wyssen

Einmal Algorithmus drüber, und schon gelingt's

In der Forschung wird KI schon seit langem angewandt. Sechs laufende Projekte, in der sie ganz besondere Fortschritte ermöglicht.

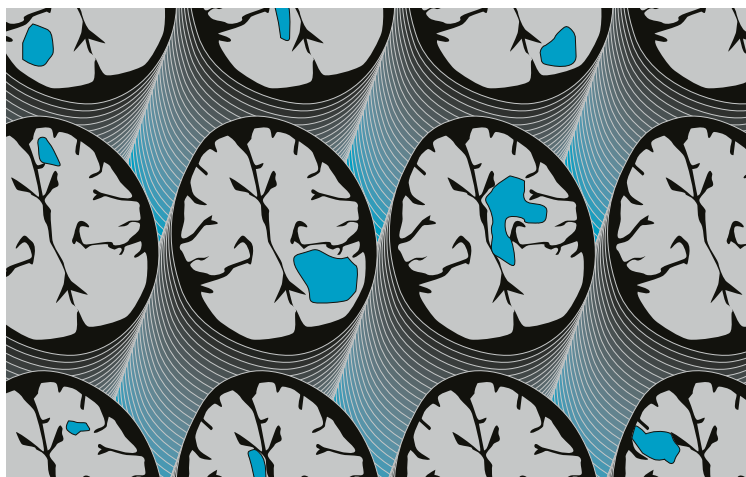
Text Hubert Filser Illustrationen Anna Haas

Dem bösartigen Tumor den Garaus machen

Mauricio Reyes, Artorg Center, Universität Bern

Technik: Bildauswertung zur Diagnose

Schnell wachsende Hirntumoren zu behandeln erfordert rasche Entscheidungen. Dafür verlassen sich die Ärzte auf MRT-Bilder. Mauricio Reyes vom Artorg Center der Universität Bern erforscht seit zehn Jahren das besonders aggressive Glioblastom. Seit verganginem Jahr setzt er für die Voraussage des Wachstums auf ein KI-System, das verschiedene Bestandteile des Tumors erkennen kann. Es soll anhand der Aufnahmen aggressive von weniger aktiven Glioblastomen unterscheiden können. Im ersten Fall liegt die Überlebensprognose bei durchschnittlich 16 Monaten, im zweiten bei mehreren Jahren. Diese Unterscheidung ist also ein zentraler Faktor für Therapieentscheidungen. «KI-Systeme helfen vor allem, Zeit zu gewinnen», sagt Reyes.



Beim Training dieser KI ging es zunächst darum, dass das System lernt, den Tumor auf den Scans in Bereiche zu unterteilen: in abgestorbene Regionen, in aktive Zonen, in denen sich die Krebszellen rasch vermehren, und in Randbereiche, in denen entzündliche Prozesse dominieren. Aus der Entwicklung dieser Zonen sollte das System lernen, unterschiedlich aggressive Varianten zu erkennen. Mit Erfolg: Die KI konnte Tumorvarianten mit einer Genauigkeit von 88 Prozent vorher-sagen. «Allerdings wussten wir nicht, auf welcher Basis die KI ihre Entscheidungen traf», erklärt Reyes. «Wir haben es oft mit Blackbox-Systemen zu tun.»

Also wollte Reyes verstehen, welche Pixel die KI verwendete. «Wir fanden heraus, dass sie vor allem auf die Randbereiche des Tumors schaute, nicht aber auf die aktiven Zonen. Das war falsch.» Shortcut-Learning nennen Forschende das Phänomen. Dabei liefern KI-Systeme

zwar halbwegs richtige Aussagen, aber aus den falschen Gründen. Sich blind auf Blackbox-Systeme zu verlassen, kann gerade in der Medizin fatale Folgen haben. Reyes schloss nun die Randbereiche des Tumors für die Bewertung aus. Sie haben nämlich genau wie abgestorbene Regionen kaum einen Informationsgehalt. Die Vorhersagegenauigkeit der KI stieg daraufhin auf 98 Prozent. Der Berner Forscher hält es für dringend notwendig, künftig KI zu entwickeln, die die Qualität von KI kontrollieren kann. Sein nächstes Projekt lautet daher «KI für KI». Bis-her gebe es so etwas nicht.

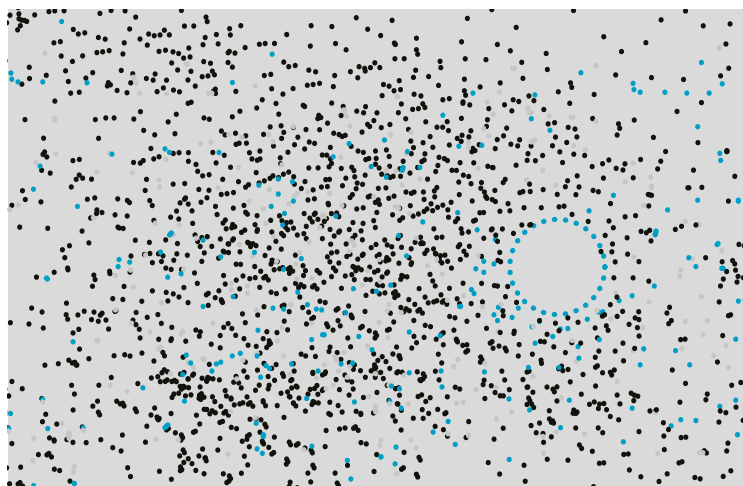
Im Rausch einer möglichen neuen Physik

Steven Schramm, Universität Genf

Technik: Suche nach Anomalien in grossen Datenmengen

Mit der Entdeckung des Higgs-Bosons am Large Hadron Collider (LHC) des Cern in Genf war das Standardmodell der Physik vollständig be-legt. Wichtige Teilkräfte wie die starke, die schwache und die elektro-magnetische Wechselwirkung waren verstanden. Die Beschreibung des Universums bleibt allerdings unvollständig, da sich Phänomene wie die dunkle Energie und die dunkle Materie damit nicht erklären lassen. Sehr wahrscheinlich existieren also weitere Teilchen.

Forschende um Steven Schramm von der Universität Genf glauben Hinweise auf unentdeckte Teilchen in den reichlich am LHC produzierten Kollisionen im Niedrigenergiebereich finden zu können. Sie wollen das sogenannte Rauschen der Daten untersuchen, das etwa durch sich überlagernde Signale, Messungenauigkeiten und Fehler der Geräte entsteht. Dieses soll mit Hilfe von KI in echte Daten verwandelt werden, um darin Muster zu erkennen. Ein solches Muster könnte auf ein Teilchen hindeuten, das den Ursprung der dunklen Materie erklärt.



«Wir hoffen, eine neue Physik zu finden», sagt Schramm. Nur mit Hilfe der KI sei es möglich, die Riesenmengen an Daten aus den LHC-Detektoren zu verarbeiten. Insgesamt ist die Aufgabe anspruchsvoll, geht es doch darum, ein 3D-Bild der elektrischen Signale aus einer Teilchenkollision zu verstehen.

Die Forschenden verwenden dafür sogenannte Graph-neuronale Netze. Trainiert wird das System mit Simulationen aus dem Standardmodell der Teilchenphysik. Damit lernt es, normale Ereignisse zu erkennen. Danach soll die KI im realen Kollisionsrauschen Abweichungen finden, also Hinweise auf eine eventuelle neue Physik. Gleichzeitig muss sie erwartete Anomalien, also Fehler eines LHC-Detektors, ausschliessen.

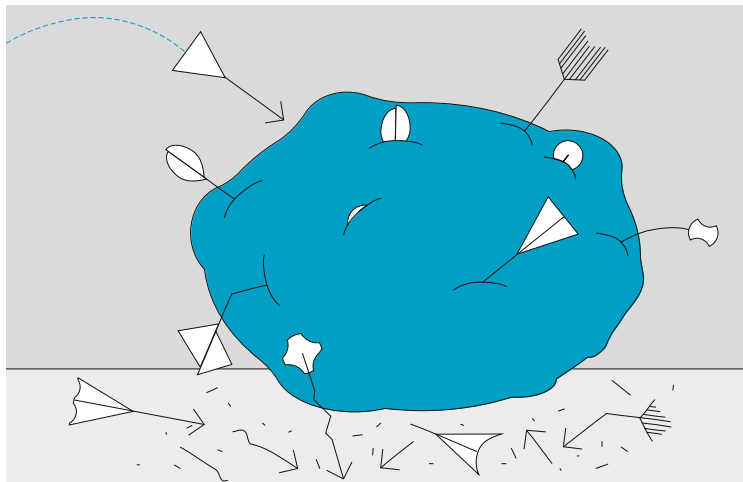
«Die Suche nach Anomalien mit Hilfe der KI ist ein völlig neuer Ansatz», erklärt Schramm. «Er ermöglicht, effizient nach Phänomenen zu suchen, die von der Theorie nicht vorhergesagt wurden.» Und wenn die Suche fehlschlägt? Dann habe man zumindest schneller Sicherheit, dass sich in den untersuchten Energiebereichen keine neue Physik verberge.

Damit das Medikament besser in die Zelle dringt

Gianvito Grasso, IDSIA Lugano

Technik: Mustererkennung zur Vorhersage von Eigenschaften

Eine der grössten Herausforderungen bei der Entwicklung neuartiger Arzneimittel ist es, die therapeutisch wirksamen Moleküle zielgerichtet an ihren Bestimmungsort zu bringen, vor allem wenn die Zielstrukturen im Inneren von Zellen liegen. Jede Zelle ist von einer Grenzschicht umgeben, einer schützenden Biomembran, die zugleich den Kontakt zur Aussenwelt vermittelt. Es ist daher enorm wichtig, das Verhalten sogenannter zelldurchdringender Peptide zu verstehen. Die kurzen Aminosäuresequenzen können die Zellmembran überwinden und Moleküle in die Zellen schleusen. «Das hat grosses Potenzial für künftige gezielte Arzneimittelgabe», sagt Gianvito Grasso vom IDSIA.



Da herkömmliche Methoden, das Verhalten dieser Peptide vorherzusagen, entweder sehr rechenaufwendige Modelle erforderten oder experimentell anspruchsvoll waren, setzt das Team am IDSIA auf künstliche Intelligenz. Algorithmen lernen, bestehende grosse Datenbanken mit bekannten zelldurchdringenden Peptiden zu analysieren und neue Peptide mit ähnlichen Eigenschaften zu identifizieren.

Trainiert wurde das System mit Daten aus der wissenschaftlichen Literatur zu zelldurchdringenden und anderen Peptiden. Diese enthalten Informationen wie die Peptidsequenz, biophysikalische Eigenschaften und ihre Fähigkeit, Zellmembranen zu durchdringen. Das

System für maschinelles Lernen des IDSIA nutzte dann diesen Datensatz, indem es die Aminosäuresequenzen bekannter Peptide beider Arten analysierte, um Strukturen zu identifizieren, die mit dem leichteren Eindringen der Peptide in die Zellen in Verbindung stehen.

Ziel war es, die Fähigkeit verschiedener Peptide, Zellmembranen zu durchdringen, genau vorherzusagen. Der Algorithmus kann aber nicht nur das Verhalten von Peptiden vorhersagen, sondern schlägt auch Modifikationen an der Peptidsequenz vor, damit diese leichter in eine Zelle eindringen, wie Grasso erklärt. Das Tool soll helfen, Peptide vor deren Synthese einfacher einzuschätzen.

Uralte Handschriften unter digitaler Lupe

Tobias Hodel, Universität Bern, und **Lucas Burkart**, Universität Basel

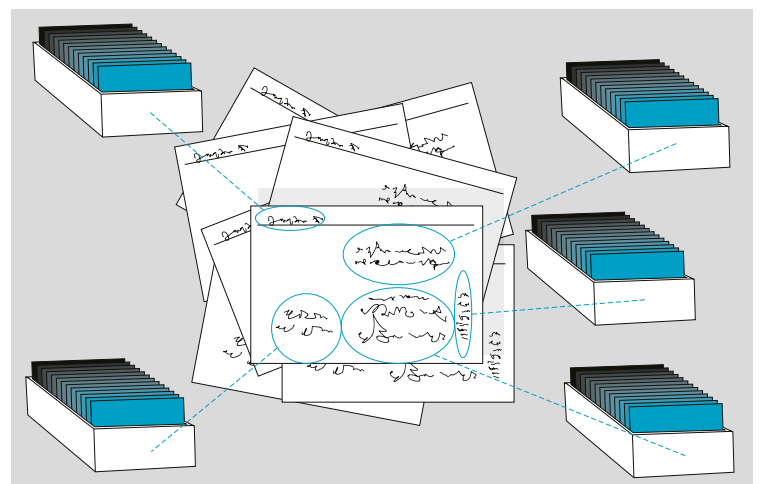
Technik: Entziffern und Auswerten alter Karteikarten

Das Historische Grundbuch der Stadt Basel umfasst mehr als 220 000 Karteikarten, mit Informationen zur städtischen Bausubstanz, die bis ins 12. Jahrhundert zurückreichen. «Für Geschichtsforschende ist so ein Bestand von immensem Wert», sagt Lucas Burkart von der Universität Basel. Allerdings sind die Notizen handgeschrieben. Um die Karteikarten für historische Forschungsfragen zugänglicher zu machen, setzen Burkart und sein Kollege Tobias Hodel von der Universität Bern nun auf KI-basierte Erschliessungsmethoden.

Ziel ist es, die Karten zu erfassen und für einen Zeitraum von 300 Jahren (1400–1700) auszuwerten. So wollen die Forschenden alle Begriffe, Vorgaben und üblichen Vorgehensweisen im Umgang mit dem städtischen Grundbesitz in Basel verstehen und damit auch den Wandel im Immobilienhandel analysieren, sowohl in Einzelfällen wie etwa dem Kloster Klingental als auch für den gesamten Stadtraum.

Die Forschenden nutzen dafür verschiedene Verfahren des maschinellen Lernens wie etwa sprachliche Datenverarbeitung. Zunächst sortiert das System Textregionen auf der Registerkarte räumlich. Im zweiten Schritt geht es darum, die Handschriften selbst zu lesen, also die Worte in der jeweiligen Sprache und den Schriftarten (Latein, alte deutsche Formen) korrekt zu erfassen. Der dritte Schritt ist der aufwendigste: «Hier zerlegen wir jede Karte in sogenannte Events», sagt Hodel. Unter einem Event versteht man dabei eine Handlung wie ein Kauf, ein Verkauf oder eine Pfändung, an der jeweils bestimmte Personen oder Institutionen wie Klöster beteiligt seien.

«Solche Events sind nicht immer eindeutig und entsprechen bisweilen einer historischen Eigenlogik, die nicht unmittelbar verständlich ist», sagt Burkart. «Hinter einem Begriff wie Zins etwa verbergen



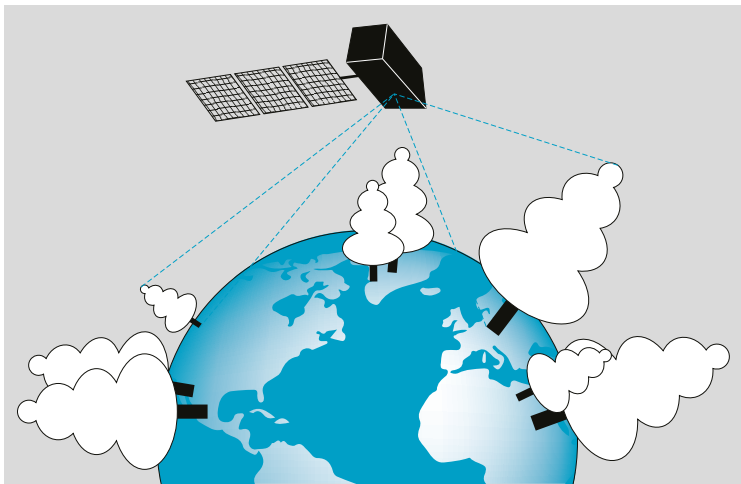
sich verschiedene Formen ökonomischer Transaktionen.» Das Projekt verfolgt auch ein übergeordnetes Ziel. «Wir versuchen, die Fehlerquote bei einzelnen Schritten der maschinellen Auswertung, etwa der Erfassung der Handschriften, zu messen», erklärt Hodel. Es gehe dabei generell um die Belastbarkeit der mit KI-Methoden gewonnenen Resultate. Solche Analysen des Systems müssten «Teil der methodischen Weiterentwicklung der Geschichtswissenschaften werden».

Vor lauter Bäumen den Klimawandel sehen

Konrad Schindler, ETH Zürich

Technik: Interpretation von Satellitenbildern

Die Höhe eines Baums zu vermessen, erscheint auf den ersten Blick keine Aufgabe, für die es eine künstliche Intelligenz bräuchte. Aber Konrad Schindler will auch nicht einen, sondern alle Bäume der Erde erfassen. Solche Daten sind von extrem hoher Bedeutung, denn sie erlauben es, mit Hilfe von Modellen die Verteilung der Biomasse auf der Erde zu bestimmen und damit auch der in der Vegetation gespeicherte Kohlenstoff.



«Wir sind dabei die Kartenmacher für die Vegetationshöhen», sagt der Zürcher Geodät und Computerwissenschaftler. Ihn interessiert, wie man mit Hilfe von KI-Methoden umweltrelevante Grössen und Zusammenhänge wie etwa die Vegetationsdichte automatisiert kartieren kann, für die das bislang nicht so gut möglich war. «Unsere Technologie erlaubt es, sogar hohe Bäume korrekt zu vermessen, die hinsichtlich Biodiversität und Biomasse besonders wichtig sind», sagt Schindler.

Die Methode ist aufwendig, denn Schindler verwendet für sein System als Trainingsdaten Pixelbilder von Kameras der beiden ESA-Satelliten Sentinel 2A und 2B. Als Referenzdaten für das Training dienen Messwerte von Baumhöhen des Laserscanners GEDI, den die Nasa an die Weltraumstation ISS montiert hat. Von dort wurden bislang Hunderte Millionen Einzelmessungen gemacht.

Da der Zusammenhang zwischen den Informationen in den Pixeln und Baumhöhen zu kompliziert ist, um daraus ein physikalisches Modell erstellen zu können, setzt Schindler auf maschinelles Lernen. Das System basiert auf einem sogenannten tiefen neuronalen Netz, das auch in der Lage ist, die statistische Unsicherheit der Ergebnisse, also eine Art Ungenauigkeit, anzugeben. Forschende nennen das «probabilistic deep learning».

Die Karte könnte in Zukunft wichtige Hinweise für geeignete Massnahmen im Kampf gegen den Klimawandel und das Artensterben lie-

fern. Sie liesse sich etwa für das Monitoring bestimmter Waldregionen oder Schutzgebiete einsetzen. Hierzulande laufen bereits Gespräche mit der Bundesforschungsanstalt für Wald, Schnee und Landschaft.

Ordnung ins Chaos der Gesundheitsdaten

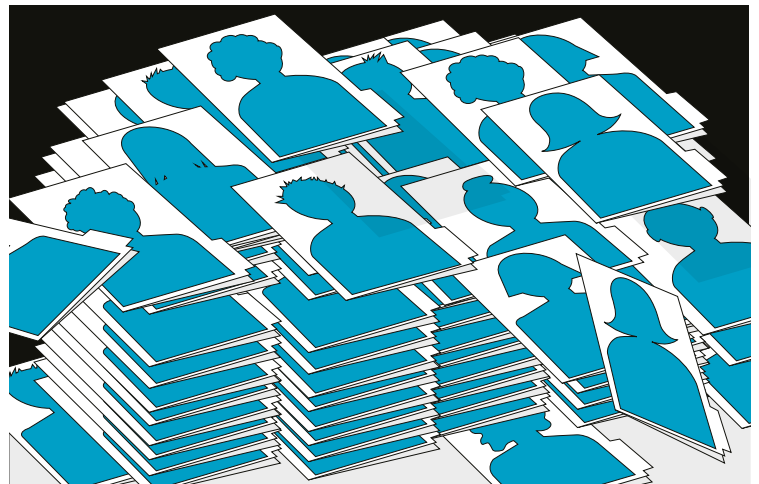
Fabio Rinaldi, IDSIA Lugano

Technik: Suche nach Ereignissen in Patientendossiers

Ältere Menschen leiden häufiger unter Nebenwirkungen von Medikamenten und sind anfälliger für unerwünschte Wechselwirkungen verschiedener Wirkstoffe. Ein Team des IDSIA in Lugano entwickelte daher eine KI, die automatisiert elektronische Patientendossiers älterer Menschen in Spitälern analysiert, um mögliche Komplikationen früh zu entdecken.

Die verwendeten Daten sind für eine KI komplex. Patientendossiers enthalten nämlich oft grosse Textmengen, in der Schweiz notieren Ärzte ihre Beobachtungen und Diagnosen zudem in den verschiedenen Landessprachen. Die Daten aus den Akten sind ausserdem sehr unstrukturiert. Um relevante Informationen daraus zu extrahieren, setzte das IDSIA-Team um Fabio Rinaldi daher sowohl auf die Technik der sprachlichen Datenverarbeitung, die auch in aktuell intensiv diskutierten Systemen wie Chat-GPT zum Einsatz kommt, als auch auf sogenanntes Text-Mining. Ziel war, dass die KI unerwünschte Ereignisse und ihre Auslöser erkennt.

Konkret ging es im Projekt um die Nebenwirkungen von Anti-Thrombose-Mitteln. Zur Qualitätskontrolle nutzte die KI Erkenntnisse zu Wechselwirkungen verschiedener Medikamente aus vier an der Studie beteiligten Schweizer Spitälern.



Ein wichtiges Thema beim Umgang mit Trainingsdaten von Patientinnen und Patienten ist die Anonymität. Das IDSIA-Team musste zunächst sämtliche Patientendossiers mit Hilfe geeigneter Tools der sprachlichen Datenverarbeitung anonymisieren. Diese Anonymisierung klinischer Berichte habe grosses Potenzial, sagt Rinaldi, auch über sein eigenes Projekt hinaus. «Wir wollten wertvolle medizinische Informationen für wissenschaftliche Zwecke nutzbar machen.» Generell könnten gesundheitsrelevante Daten, wenn sie in grossem Umfang zusammengeführt werden, medizinische Erkenntnisse zutage fördern, die bei einer isolierten Betrachtung von nur wenigen Fällen nicht sichtbar seien.

Hubert Filsler ist Wissenschaftsjournalist in München.

Erwarten wir das Unerwartete

Eine Kombination von riesigen Sprachmodellen wie GPT und dem bestärkenden Lernen könnte dazu führen, dass künstliche Intelligenz auch die Forschung selbst fundamental verändert. Ein Blick in die Kristallkugel.

Text Roland Fischer

Die neuesten Modelle der künstlichen Intelligenz (KI) bringen Umwälzungen in mythischen Grössenordnungen. So verglich Thomas Friedman, bekannter Kolumnist der New York Times, die gegenwärtige Situation mit Sprachgeneratoren wie GPT-4 und Bildgeneratoren wie Stable Diffusion mit dem Einfluss von Gott Prometheus, der in der Sage den Menschen Feuer brachte und so die Zivilisation startete. Alles Bisherige würde über den Haufen geworfen. «Wir können nicht kleine Anpassungen vornehmen, sondern müssen alles ändern: wie wir kreativ sind, miteinander konkurrieren, zusammenarbeiten, lernen, regieren und, ja, wie wir betrügen, Verbrechen begehen und Kriege führen.»

Forschen kommt in Friedmans Liste nicht vor. Ist es, weil das Erkunden noch nicht kartierter Landschaften die Domäne des Menschen ist? Diese Frage beschäftigt auch den EPFL-Professor Rüdiger L. Urbanke, der beim Geneva Science and Diplomacy Anticipator (GESDA) versucht, das Zukunftspotenzial von KI abzuschätzen: «Wenn Sie mich vor drei Monaten gefragt hätten, welche Jobs am ehesten bedroht sind durch KI, hätte ich wohl noch das genaue Gegenteil gesagt, dass die Jobs nämlich überall da sicher sind, wo ein gewisses Mass an Originalität wichtig ist. Doch nun, mit GPT und Stable Diffusion, kann man sich plötzlich vorstellen, dass es die ganze Kreativindustrie mit einer ganz ordentlich arbeitenden KI-Konkurrenz zu tun bekommt.»

KI faltet erfolgreich Proteine

Dass KI einmal eigenständig forschen und Neues entdecken wird, davon geht jedenfalls Pedro Domingo, emeritierter Professor der University of Washington, aus: «Manche sagen, maschinelles Lernen könne statistische Regularitäten in Daten finden, aber niemals etwas Tiefgründiges wie Newtons Gravitationsgesetz entdecken», schrieb er 2016 in seinem erfolgreichen Einführungsbuch für KI, «The Master Algorithm»: «Zweifellos ist dies bisher nicht passiert, aber ich wette, dass es einmal so weit sein wird.» Die Frage kam damals im

Zusammenhang mit Big Data auf. Sie veranlasste auch den Tech-Autor Chris Anderson, im Magazin Wired ein Manifest einer neuen Wissenschaft zu veröffentlichen: «Korrelation macht Kausalität überflüssig, und wissenschaftlicher Fortschritt ist ohne kohärente Modelle, Einheitstheorie oder mechanistische Erklärung möglich. Wir müssen nicht am Alten festhalten. Es ist Zeit zu fragen: Was kann die Wissenschaft von Google lernen?»

Die gegenwärtigen Umwälzungen waren damals höchstens zu erraten. Doch inzwischen hat sich KI als Werkzeug in vielen Feldern der akademischen Forschung etabliert, ohne gleich erkenntnistheoretische Grundsatzfragen aufzuwerfen. Als jüngstes Beispiel gilt AlphaFold, ein Programm, das ein altes Problem in vielen Fällen recht gut lösen kann: die Faltung eines Proteins präzise vorherzusagen. Für die Überraschung sorgte die Firma DeepMind, die ihre Methode beim Go-Spiel perfektioniert hatte. Dabei erhielt die KI keine genaue Anleitung zur Lösung des Problems. Sie wurde optimiert, indem ihre Fortschritte belohnt und ihre Rückschritte bestraft wurden – mit Punktegutschriften und -abzügen. Bei der Methode des sogenannten bestärkenden Lernens wird der KI überlassen, wie genau eine Lösung am Ende zustande kommt. So entstehen oft auch unerwartete Zugänge und Resultate, die Go-Spezialisten damals mit der Entdeckung eines neuen Kontinents verglichen hatten.

Ob Forschende dereinst dank kreativer KI auf die grossen weissen Flecken der Landkarte vorstossen können, bleibt eine offene Frage. Jack Clark, KI-Evaluationsspezialist und Autor des einflussreichen Newsletters Import AI, sieht jedenfalls Parallelen in der Entwicklung des bestärkenden Lernens heute und der grossen Sprachmodelle wie GPT vor ein paar Jahren. Es lasse sich eine Gesetzmässigkeit zwischen Grösse und Leistung der Modelle erkennen. Sobald voraussehbar sei, dass sich die Leistung der Systeme mit dem Einsatz zusätzlicher Ressourcen entscheidend steigern lässt, werde sich die Entwicklung zusätzlich beschleunigen. Das nächste grosse Ding könn-

te folglich die Verbindung von bestärkendem Lernen mit den Sprachmodellen sein: Vielleicht lernen die Forschenden Neues dann nicht mehr mit KI-Methoden, sondern ganz direkt im Gespräch mit der KI. Vielleicht ist das Zeitalter der methodischen Forschung doch noch lange nicht vorbei, wie Anderson meinte, doch die Modelle, Theorien und mechanistischen Erklärungen werden neu von der KI erdacht – nicht mehr von Menschen.

Mehrjahresplan unmöglich

Diese Vision mag verstiegen klingen. Zwar stimmt es, dass die grossen Sprachmodelle «nur» die plausibelste Fortsetzung einer Spracheingabe berechnen. Aber die Resultate simpel beiseitezurück und zu argumentieren, dass die Modelle nur so tun, als wüssten sie etwas, ohne anwendbares Wissen zu erlangen, das funktioniert spätestens seit GPT-4 nicht mehr. Zur Veranschaulichung: Forderte man das Vorgängermodell GPT-3 zu einer Schachpartie auf, spielte es mit grosser Überzeugung Blödsinn. Doch mit dem nächstgrösseren Modell ist ein wichtiger Schritt passiert: Die KI hat nun auch die Regeln des Spiels gelernt und hat dabei ein durchaus respektables Niveau erlangt.

Urbanke von GESDA plädiert jedenfalls dafür, das Unerwartete zu erwarten. Es gäbe immer wieder Überraschungen im weiten Feld der KI. «Man kann keinen Zehnjahresplan machen, aber man muss nahe an den Entwicklungen bleiben, und das heisst, man muss Grundlagenforschung betreiben.» Dabei stellt sich natürlich wieder die Frage nach den Ressourcen, wenn die Modelle immer mächtiger werden. Können nationale Infrastrukturen überhaupt mit den grossen Firmen mithalten bei der Weiterentwicklung der KI? Noch einmal Urbanke: «Es muss nicht ein Cern sein. Ein paar Player, die zusammenspannen, reichen schon.» Er sei jedenfalls nicht besorgt diesbezüglich: «Wir sind gut aufgestellt in der Schweiz.»

Roland Fischer ist Wissenschaftsjournalist in Basel.

Das meistverehrte Lächeln der Welt hat eine Appropriation durch moderne Technologie erfahren. KI und Künstler haben Leonardo da Vincis Mona Lisa in einen Roboter verwandelt. Lächeln inklusive.

Bild: Jonas Wyssen

