

Künstliche Intelligenz

Philosophische Tage beschäftigten sich mit vielseitigem Thema

Ein hochaktuelles Thema war Gegenstand der Philosophischen Tage, die vom 5. bis 7. Oktober 2023 in der Katholischen Akademie in Bayern stattfanden: *Künstliche Intelligenz*. Die Künstliche Intelligenz, kurz KI genannt, spielt eine immer wichtigere Rolle in vielen gesellschaftlichen Bereichen. So ist die KI auch Gegenstand philosophischer Debatten geworden. Dabei geht es beispielsweise um ein philosophisches Verständnis der neuen Technologie und die Auslotung des Verhältnisses von Mensch und Maschine.

Das Themenspektrum der Referate an den Philosophischen Tagen reichte von der historischen Entwicklung von KI bis hin zum Einsatz von KI in Unternehmen und dem Umgang mit KI in Kontexten von Kunst und Kreativität. Nachfolgend dokumentieren wir den Vortrag von Dr. Rudolf Seising zur Geschichte der KI sowie das Referat von Prof. Dr. Karoline Reinhardt, die darauf schaut, wie die KI unsere demokratische Gesellschaft, etwa unser Verständnis von Öffentlichkeit und öffentlicher Kommunikation, beeinflusst.

Wie viel KI brauchen wir? Wie viel wollen wir?

von Johannes Schießl

Künstliche Intelligenz (KI) spielt heute in vielen gesellschaftlichen Bereichen eine immer größere Rolle: angefangen von autonom fahrenden Autos über medizintechnische Verfahren bis hin zu Kunstwerken, die von künstlicher Intelligenz kreiert werden. In all diesen Feldern prägen die neuen Technologien das Zusammenleben der Menschen, verändern Lernprozesse oder auch unsere Wahrnehmung von der Wirklichkeit als solcher. Manche sprechen sogar davon, dass KI-Systeme (irgendwann) auch ein Bewusstsein oder eine Seele haben könnten. Wie Menschen sprechen, denken oder singen können sie jedenfalls bereits.

Auch die Philosophie widmet sich diesen Debatten intensiv. Dabei geht es beispielsweise um ein philosophisches Verständnis dieser neuen Technologie

Keine Maschine wird moralisch handeln können, da sie über kein Bewusstsein verfügt. Jeder Computer setzt letztlich auf Binarität, während die Philosophie nicht bei Null und Eins stehen bleiben kann.

oder auch um erkenntnistheoretische Fragen ihrer Erklärbarkeit. Immer wieder wird dabei, auch vor einem historischen Hintergrund, die Differenz zwischen Mensch und Maschine betont. Praktische Philosoph:innen wiederum diskutieren, wie genau dieses Verhältnis aber dann zu deuten ist, ob man sinnvollerweise vom Vertrauen in KI sprechen kann oder wie neue Praktiken an der Schnittstelle zwischen menschlichem Handeln und KI beschrieben und bewertet werden können. Die Philosophischen Tage 2023 haben genau diese Fragen aufgegriffen und vor allem die praktisch-philosophische Dimension im Umgang mit KI-Systemen beleuchtet.

Eine für die Entwicklung der KI entscheidende Antwort hat Alan Turing im 20. Jahrhundert gegeben, indem er Denken mit Rechnen gleichsetzte und prognostizierte, dass Maschinen einst denken würden.

Professor Michael Reder von der Hochschule für Philosophie in München, der die Tage bereits zum vierten Mal geleitet hat, stellte in seiner Einführung die Grundfrage so: „Können KI-Systeme etwas, was eigentlich genuin menschlich ist?“ Um gleich hinzuzufügen: „Keine Maschine wird moralisch handeln können“, da sie über kein Bewusstsein verfüge. Jeder Computer setze letztlich auf Binarität, während die Philosophie nicht bei Null und Eins stehen bleiben könne. Seit ungefähr einem Jahr habe die Debatte um KI ungeheuer an Fahrt aufgenommen, Aufgabe der Philosophie sei es, Anfragen zu formulieren, wie etwa: Wieviel KI brauchen wir? Wieviel wollen wir? Wieviel haben wir schon? Und schließlich: Wer übernimmt die Verantwortung?

Den Auftaktvortrag unter der Überschrift *KI und Autonomie* hielt Professor Benjamin Rathgeber, der Natur- und Technikphilosophie ebenfalls an der Münchner Hochschule für Philosophie lehrt. Durch die Feuilletons geisterten viele metaphorische Schlagworte, wirklich neu an der KI sei, dass eine Technologie nicht bloß als Mittel fungiere, sondern selbst auch Zwecke setze. Die vier wichtigsten Anwendungsbereiche seien die autonome Mobilität, die humanoide Robotik, eine sich selbst steuernde Produktion und KI-basierte Forschung. Rathgeber macht sechs „Aufstufungen“ der technologischen Entwicklung aus: die Instrumentalisierung, die Maschinisierung, die Reproduzierung, die Kybernetisierung, die Digitalisierung und schließlich die „Kognisierung“. Auf jeder Stufe gehe es um eine weitere Verselbstständigung, Impulse von außen wür-

den zunehmend in das Innere des Prozesses verlagert. Das führe zu der Frage, ob eine zunehmende „Autonomisierung“ von Technologie die Aufgabe der eigenen Autonomie zur Folge habe.

Dazu sei eine Klärung des Begriffs der Autonomie notwendig, der ursprünglich keine individuelle Kategorie gewesen, sondern in der griechischen Antike eher auf die Polis bezogen gebraucht worden sei. Erst in der Neuzeit, vor allem in Kants *Grundlegung zur Metaphysik der Sitten*, sei der Begriff auf die einzelne Person übertragen worden. Benjamin Rathgeber zerlegte den Begriff *Autonomie* in seine Bestandteile. Es mache einen großen Unterschied, ob der erste Bestandteil des Worts,

Die vielen Geschichten der KI lautete der Titel des Referats (siehe [Seite 46 bis 51](#)) von Rudolf Seising, der als wissenschaftlicher Mitarbeiter im Forschungsinstitut für Technik- und Wissenschaftsgeschichte am Deutschen Museum arbeitet. „Maschinen denken nicht“, formulierte Seising pointiert, „das tun nur Lebewesen – und da nicht alle“. Platon habe vor aller Naturwissenschaft Denken als inneres Gespräch der Seele mit sich selbst verstanden. Descartes habe den Ort des Denkens in der Zirbeldrüse verortet, da sie als einziges Teil im menschlichen Gehirn unpaarig auf-trete. Das sei zwar heute nicht mehr haltbar, aber doch ein früher Versuch gewesen, die Verbindung von Körper und Geist zu denken.



Foto: alicia Shepeleva / Shutterstock.com

Manche sprechen davon, dass KI-Systeme (irgendwann) auch ein Bewusstsein oder eine Seele haben könnten. Wie Menschen sprechen, denken oder singen können sie jedenfalls bereits.

nämlich das griechische *autos*, nun „von selbst“, „an sich selbst“, „auf sich selbst“, „über sich selbst“ oder „um seiner selbst“ meine. Auch *no-mos*, der zweite Bestandteil des Worts, sei vieldeutig, er könne Regel, Verwirklichung, Bestimmung, Gesetz oder Zweck bedeuten. Vieles, was unter Autonomie laufe, meine eher eine technische Automatisierung. Es bleibe aber das Problem, dass sich bei einer vollständigen Reduktion von Autonomie auf die Automatisierung die Autonomie auflöse.

Eine für die Entwicklung der KI entscheidende Antwort habe Alan Turing im 20. Jahrhundert gegeben, indem er Denken mit Rechnen gleichsetzte und prognostizierte, dass Maschinen einst denken würden. Ab der Mitte des letzten Jahrhunderts habe es interdisziplinäre Konferenzen gegeben, die der Frage nachgingen, was im Gehirn eigentlich passiere. Dabei wären zunehmend Parallelen zur Welt der Computer gezogen worden. Der Begriff *Artificial Intelligence* sei erstmals im Jahr 1956 gefallen. Als eine

ihrer wichtigsten Fähigkeiten müsse man die Mustererkennung nennen, die auch in der heutigen Forschung zu neuronalen Netzen entscheidend sei. Ein Beispiel für maschinelles Lernen bilde der Schachcomputer. „Ein Mensch kann neun Züge vorausdenken“, so Seising, „aber nur, wenn er sehr gut ist“. Ein Computer übertreffe das bei Weitem, zudem speichere er immer mehr erfolgreiche Spielzüge. 1997 dann habe der Schachcomputer *Deep Blue* den Weltmeister Garri Kasparow geschlagen.

Karoline Reinhardt, Juniorprofessorin für angewandte Ethik an der Universität Passau, ging in ihrem Beitrag (siehe [Seite 52 bis 56](#)) anhand des seit einem Jahr heftig diskutierten ChatGPT auf ethische Probleme aktueller KI-Entwicklungen ein und zeigte

Technologie ist nicht neutral, sie ist gerichtet und existiert nicht in einem Vakuum. KI kann also nicht als bloßes Werkzeug betrachtet werden, wie häufig behauptet.

Gefahren für demokratische Gesellschaften auf. Large Language Models basierten auf Mustererkennung und -reproduktion bzw. -produktion. Buchstabensequenzen würden in Zahlen transformiert, die Wahrscheinlichkeit von Wortfolgen würde berechnet, die dann automatisch wieder in Wörter übersetzt würden, wobei das Verfahren ständig optimiert werde. Basierend auf riesigen Datenmengen kämen „wahrscheinliche, nicht wahre“ Ergebnisse zustande.

Der häufig geäußerten Behauptung „Es ist bloß ein Werkzeug, wir müssen nur lernen, mit ihm umzugehen“ setzte Karoline Reinhardt entgegen: „Technologie ist nicht neutral, sie ist gerichtet und existiert nicht in einem Vakuum.“ Technologie verfüge immer über intrinsische Eigenschaften,



Im Viereckhof tagte der Arbeitskreis mit Prof. Dr. Christian Au, Professor für Allgemeine Betriebswirtschaftslehre an der Hochschule Mainz. Hier ging es um Chancen und Grenzen von KI im Arbeitsalltag von Unternehmen.

die bestimmte Nutzungsweisen ermöglichten und andere ausschlossen. Zudem seien Technologien zweckbestimmt, diese Zwecke seien „nicht naturläufig, sie werden gesetzt und können deswegen ethisch bewertet werden“. Schließlich seien Technologien immer in menschliche Praktiken eingebunden und prägten sie wie auch das menschliche Selbstbild und Selbstverhältnis nachhaltig. Sprachmodelle seien nie wertneutral und nähmen Einfluss auf die Rahmenbedingungen der Technikgestaltung und -regulierung. So drohe eine Fragmentierung von öffentlicher Kommunikation und eine Unterminderung von Normen. Auch könne die Verwendung von generativen Sprachmodellen zu Misstrauen gegenüber der Medienkommunikation und zu öffentlicher Zurückhaltung führen. Gesellschaftliche Debatten über KI seien nur möglich, wenn Transparenz über deren Einsatz und Funktionsweise herrsche.

Über den Einsatz von KI in Unternehmen sprach Professor Christian Au von der Hochschule Mainz. Er stellte sich zunächst die Frage, was neu sei an der aktuellen KI-Welle. Die Entwicklung der Technologie sei seit ihren Anfängen in Wellen verlaufen, auf mehrere „Winter“ seien immer wieder neue Aufbrüche gefolgt. Wenn man sich nun das Suchverhalten auf *Google Trends* ansehe, könne man feststellen, dass die Nachfragen seit Ende 2022 exponentiell gestiegen seien, sie hätten sich nahezu verfünffacht. Beim Job-Portal *LinkedIn* seien

KI-Kompetenzen doppelt bis dreifach stärker nachgefragt gewesen. „Es denkt“ sei zwar nach wie vor falsch, so Christian Au, aber „es funktioniert“, vor allem weil es inzwischen einfach in der Handhabung sei. Gerade bei der Erkennung von Handschrift, von Sprache und von Bildern habe die KI ungeheuer schnell gegenüber den menschlichen Fähigkeiten aufgeholt und sie inzwischen sogar überholt. Das erzeuge „hektische Betriebsamkeit in den Unternehmen“.

Aber welchen ökonomischen Mehrwert könne die KI nun stiften? Sie sei in der Lage, viele Prozesse zu automatisieren, nunmehr auch bei den „Schreibtisch-Tätigkeiten“ und sogar bei Führungsaufgaben, etwa wenn die Auswahl von Dokumenten und relevanten Fakten gefragt sei, wenn es um das Finden, Gruppieren und Priorisieren von Fakten gehe oder um die Zusammenfassung von Ergebnissen. So könne die Arbeitszeit verkürzt werden und gleichzeitig die

Die KI ist in der Lage, viele Prozesse zu automatisieren, nunmehr auch bei den „Schreibtisch-Tätigkeiten“ und sogar bei Führungsaufgaben, etwa wenn die Auswahl von Dokumenten und relevanten Fakten gefragt ist.

Die Beurteilung von kreativer KI schwankt insgesamt zwischen deren Anthropomorphisierung und ihrer Dämonisierung. Die Aufmerksamkeit richtet sich vor allem auf die Technik, nicht auf die Kunstschaffenden. Kunst aber ist stets Reflexion und kann auch mit KI kreativ umgehen.

Qualität verbessert. Dabei stehe der Einsatz generativer KI in deutschen Unternehmen noch ganz am Anfang. Insgesamt jedoch hänge die Qualität der Ergebnisse und der Grad der Unvoreingenommenheit der Modelle entscheidend von den eingegebenen Trainingsdaten ab. „Je besser die Daten, desto besser das Modell“, resümierte Christian Au.

Mit dem Thema *Kunst, Kreativität und KI* beschäftigte sich Stephanie Catani, Professorin für Neuere deutsche Literaturgeschichte an der Universität Würzburg. Die Erstellung etwa eines Rembrandt-Gemäldes oder eines Harry-Potter-Kapitels durch KI habe für intensive Diskussionen in den Feuilletons gesorgt. Dabei sei aber immer zu bedenken, dass generative KI-Modelle synthetische Daten erzeugten, die in ihren Eigenschaften

den Originaldaten ähnelten, auf die das jeweilige KI-Modell trainiert worden sei. Generative Kunst zwingt jedoch zu einer Aufweichung zwischen Kunst und Medien. In Deutschland werde das eher kulturpessimistisch beurteilt, im Hintergrund stehe immer noch der konventionelle Geniebegriff. Der weltweit erfolgreiche KI-Künstler Refik Anadol zum Beispiel habe mit seiner auf dem Davoser Weltwirtschaftsforum gezeigten Inszenierung *Coral* auf 100 Millionen Einzelbilder zurückgegriffen, das Werk gehe aber weit über die KI hinaus, da es vom Künstler „dirigiert“ worden sei und durchaus ethische Implikationen habe.

Es stellten sich aber auch drei kritische Fragen, zunächst die nach der Urheberschaft, gerade bei Bildern. Es gehe dabei zwar nicht um Plagiate, aber die Trainingsdaten stammten doch aus realen Bildern von Urhebern, die Rechte hätten und eine Zukunft bräuchten. Das zweite Problem sei die Datenintransparenz: Welche Daten liegen den Modellen zugrunde und woher kommen sie? Die neueste Version von GPT etwa gebe das nicht mehr an. Und schließlich sei das Crowdfunding hinter der KI zu kritisieren, die ihre Datenmenge nur durch die Ausbeutung von Datensortierenden im globalen Süden schaffen könne. Die Beurteilung von kreativer KI schwanke insgesamt zwischen deren Anthropomorphisierung und ihrer Dämonisierung. Die Aufmerksamkeit richte sich vor allem auf die

Technik, nicht auf die Kunstschaffenden. Kunst aber sei stets Reflexion und könne auch mit KI kreativ umgehen. Man dürfe sie nur nicht auf den Output reduzieren, sondern müsse immer den Prozess mitdenken.

Professor Alexander Filipović, Sozialethiker an der Universität Wien, und seine ehemalige Promovendin Anna Puzio von der Universität Twente in den Niederlanden gingen zum Abschluss der Philosophischen Tage der Frage nach, inwieweit Freiheit und Autonomie als Kategorie für Mensch-Roboter-Interaktionen dienen können. Anna Puzio verwies auf die Vielfalt der bereits heute möglichen Beziehungen zu Robotern, das reiche von Sexrobotern, Freundschaftsrobotern über soziale Roboter und solche im Haushalt bis zum Einsatz von Robotern in der Industrie und in der Medizin.

Alexander Filipović unterschied verschiedene Aspekte von Autonomie, nämlich die Autonomie, Beziehungen einzugehen, in Beziehungen autonom zu bleiben und schließlich sogar die Tatsache, dass durch Beziehungen Autonomie vergrößert werden könne. Mit Axel Honneth plädierte Filipović für einen „sozialen Begriff von Freiheit“, der mehr als negative und reflexive Freiheit bedeute, da er den Begriff auch in die äußere Wirklichkeit verlege.

Anna Puzio stellte dem das Konzept von Karen Barad entgegen, die sich dem Neuen Materialismus zuordnet. Mit ihrer Kritik am Anthropozentrismus könnten auch „nicht-menschliche Entitäten“ wie Tiere oder Roboter in den Blick genommen werden. Mensch, Körper und nicht-menschliche Entitäten existierten nicht vorgängig, sondern entstünden erst in und durch Beziehungen. Tiere und Roboter bekämen so eine „eigene Agency“, die freilich die menschliche Autonomie begrenzen könnte, die nicht mehr Angelegenheit des isolierten Individuums bleibe. Eine Position, die zu intensiven Diskussionen anregte, wie auch die anderen Vorträge dieser Philosophischen Tage. ■



Bild: erstellt mithilfe von KI / canva.com



Bild: erstellt mithilfe von KI / canva.com

Es ist beeindruckend, wie die KI aus eingegebenen Schlagwörtern, den Prompts, Bilder generiert. Zur Erstellung der oben gezeigten Bilder wurden die Prompts „Buchstabe K in der Natur“ und „Buchstabe K auf Wasser“ verwendet.

Von der Mannigfaltigkeit der KI-Geschichte

Imitation menschlicher Intelligenz und statistische Voraussage prägten den Begriff der „Künstlichen Intelligenz“ von Rudolf Seising

Seitdem ChatGPT gegen Ende des Jahres 2022 für jedermann und kostenfrei nutzbar wurde, lässt sich über Künstliche Intelligenz noch aufgeregter und emotional aufgewühlter diskutieren als zuvor. Chancen und Risiken dieser Technologie sind abzuwiegen, die zu erwartenden Veränderungen in Schule, Beruf, Studium, Wissenschaft und Wirtschaft, Privatleben und Arbeitswelt zu erörtern, soziale und moralische Fragen zu stellen und zu beantworten. Aber nicht viele Gedanken werden der Herkunft von Künstlicher Intelligenz (KI) gewidmet: Woher kommt und woraus besteht sie eigentlich, worauf gründete sie sich, wie hat sie sich entwickelt bzw. wie wurde sie konstruiert? – KI ist von Beginn an mit dem Aufkommen der Computer verbunden gewesen, die schon in den Anfangszeiten „Elektronengehirne“ oder „Denkmaschinen“ genannt wurden. Auch war sie schon von Beginn an Thema von Kurzgeschichten, Romanen, Comics und Kino- bzw. Fernsehfilmen mit Zukunftsvisionen. Fähigkeiten früherer oder heutiger KI-Systeme wurden oft als Mischungen aus realen Technologieeigenschaften und fiktionalen oder gar phantastischen Kreationen geschildert. Da die Begriffe Intelligenz und Denken gar nicht einhellig definiert sind, wird meist auch deren gesamtes Bedeutungsspektrum genutzt, um plakativ und reißerisch in der Öffentlichkeit Schlagzeilen zu machen. Die Erzählungen wurden in der zweiten Hälfte des 20. Jahrhunderts zudem gern mit denen der Science-Fiction um Weltraumreisen und Atomenergie-Euphorie vermischt.

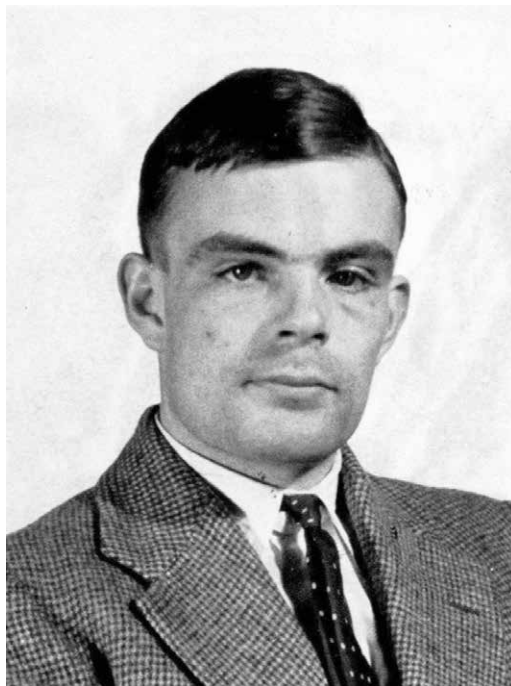
Nüchtern betrachtet ist ein anderes Bild zu zeichnen, um den KI-Entwicklungen historisch gerecht zu werden. Da KI einerseits aus computerisierten Systemen besteht, andererseits aber mit den geisteswissenschaftlichen Begriffen namentlich der Intelligenz, aber auch des Denkens und des Bewusstseins untrennbar verbunden ist, haben Entwicklungen aus beiden Wissenschaftskulturen zu dem geführt, was uns heute beschäftigt.

Computer sind logische Maschinen. Sie realisieren die logischen Verknüpfungen durch Schaltkreise, die entweder elektrischen Strom führen, oder nicht. Dieses binäre Prinzip ermöglicht die logische Wahrheit durch Ketten von Aussageverknüpfungen zu verfolgen. Logisches Denken ist aber nur ein Aspekt des Phänomens, das wir Intelligenz nennen. Viele weitere kognitive Fähigkeiten wurden in der Psychologie untersucht, um Intelligenz zu erfassen, z. B.

das Wahrnehmen und Erkennen, die Aufmerksamkeit und das Erinnern, das Problemlösen und das Lernen. Diese geistigen oder mentalen Vermögen schreiben wir den Menschen und einigen wenigen anderen Lebewesen zu. Im 19. Jahrhundert machten Psychologen die Untersuchungen dieser Fähigkeiten zu einer empirischen Wissenschaft, als sie das Erleben und Verhalten der Menschen experimentell erforschten.

Da man nach dem II. Weltkrieg damit begann, den Computern die mühsame Kärnerarbeit sehr langer logischer Schlussfolgerungsketten zu übergeben, die als der regelgeleitete Kern menschlichen Denkens angesehen wurde, war es nur folgerichtig, sich zu fragen, ob diesen Maschinen oder entsprechende Weiterentwicklungen nicht auch zu nicht-logischem Denken und anderen kognitiven Merkmalen

befähigt werden können. So formulierte im Jahre 1950 der britische Mathematiker Alan Turing in seinem Artikel *Computing Machinery and Intelligence* in der Psychologie-Zeitschrift *Mind* die Frage, ob Computer denken können. Das Vermögen zu denken bezeichnete er dabei als „intelligence“ und zur Klärung des Sachverhaltes schlug er ein „Imitation Game“ vor, bei dem ein Mensch aufgrund der Antworten, die er von einem Mitspieler auf seine Fragen bekommt, entscheiden soll, ob der antwortende Mitspieler menschlich ist oder eine Maschine. Turing kam zu dem Schluss, dass einem Computer, den ein Fragensteller für einen Menschen hält, weil seine Antworten solche, die



Der britische Mathematiker Alan Turing, hier auf einer Fotografie im Jahr 1936, ging bereits in den 1950er Jahren der Frage nach, ob Computer denken können.

Bild: Wikimedia Commons, Public Domain

Menschen geben könnten imitieren, ein dem Menschen gleichwertiges Denkvermögen zugesprochen werden müsse und er deshalb intelligent genannt werden sollte.

Die Computer dieser Zeit waren elektrotechnische Geräte, mit denen die Operationen der klassischen Aussagenlogik wie „a und b“, „c oder d“ und „wenn e, dann f“ in elektrische Schaltungen übersetzt werden konnten. Es waren Maschinen, die ihnen eingegebene Daten als Symbole verarbeiteten. Mit diesen „computing“ genannten Prozessen sollte also jegliches menschliche Denken in Maschinen nachgeahmt werden können? – Das konnten Computer aus dem Jahre 1950 nicht! Das wusste selbstverständlich auch Turing, aber er war davon überzeugt, dass zukünftige Computer das Leben der Menschen verändern würden und man um das Jahr 2000 wie selbstverständlich von „denkenden Maschinen“ sprechen würde, weil sich der Begriff „Denken“ dann gewandelt haben würde. Hier hat sich Turing geirrt: Auch im Jahr 2024 schreibt nicht jeder den KI-Systemen das Denken ganz selbstverständlich zu, wenn sich auch entsprechend unreflektierte Äußerungen immer weiter verbreiten!

Ob intelligente Maschinen konstruiert werden könnten, fragten sich zur Mitte der 1950er Jahre die Mathematiker John McCarthy und Marvin Minsky und die Elektrotechniker Claude Shannon und Nathaniel Rochester. Darüber wollten sie mit anderen Forschern bei einem Treffen im US-amerikanischen Dartmouth College diskutieren, wie sie in einen Förderantrag an die Rockefeller Foundation schrieben. Das Treffen fand 1956 statt und heute wird es oft der Beginn der Forschungen zur KI genannt. Dort stellten Herbert Simon und Alan Newell das Programm *Logic Theorist* vor, das menschliches Denken beim Beweisen logisch-mathematischer Probleme simulierte. Ihr zwei Jahre später entwickeltes Programm zur Lösung ganz allgemeiner Probleme – der *General Problem Solver* (GPS) – wurde wieder mit dem Anspruch entwickelt ganz allgemein menschliches Denken simulieren zu können. Dieses Projekt scheiterte an der in der Sache liegenden Komplexität.

Nur drei Jahre später gründeten Minsky und McCarthy am MIT die erste Forschergruppe mit dem Ziel, KI-Systeme zu bauen. 1962 ging McCarthy nach Kalifornien, um an der Universität Stanford ein weiteres KI-Laboratorium aufzubauen. Im

Jahre 1960 veröffentlichte Minsky den vielbeachteten Artikel *Steps towards AI*, in dem er zusammenfasste, welche Anforderungen an KI-Systeme seines Erachtens erfüllt werden müssen: Sie sollten in Datenvolumina gezielt suchen, Muster erkennen, lernen und planen können und auf einem Computer ablaufbar konstruiert werden. Zur Lösung solcher und ähnlicher Probleme

schrieben Wissenschaftler dann Programme, die in einem Computer ablaufen sollten. Dabei ging es auch darum, mögliche Ereignisse im Voraus zu berücksichtigen und die jeweils dann angemessene Handlungsoption in Unterprogrammen zu berücksichtigen. Für jeden Problemtyp entstanden so Routinen, in denen immer wieder gleiche Schrittfolgen von programmierten Algorithmen durchgeführt wurden, z. B. zum Beweisen mathematischer Sätze, zur Zeichenerkennung in Laut und Schrift. Zudem sammelte man für gut eingrenzbarere Anwendungsgebiete z. B. in der Geologie, der Medizin oder beim Militär Faktenmaterial, das in Wissensbasen gespeichert wurde. Durch Zusammenschluss dieser Wissensbasen mit den logischen Folgerungsmechanismen entstanden in den 1970er Jahren sogenannte Expertensysteme, etwa zur Analyse von Massenspektren, um chemische Strukturen zu identifizieren, zum Diagnostizieren von Schäden bei technischen Systemen oder auch von Krankheiten aufgrund von Symptomen und Untersuchungswerten bei den Patienten. So verfügte das medizinische Expertensystem MYCIN über 450 Regeln, die zu Vorschlägen für Therapien bei Infektionskrankheiten führten. Diese Regeln wurden in Zusammenarbeit mit Experten (in diesem Falle erfahrene Mediziner) aufgestellt und das System konnte aus eingegebenen und gespeicherten Fakten mittels der Regeln Schlüsse ziehen. Letztendlich war diese Technologie zwar begrenzt erfolgreich, enttäuschte aber gemessen an den vollmundigen Voraussagen.

GOFAI und NFAI

1985 erschien das Buch *Artificial Intelligence: The Very Idea* des Philosophen John Haugeland. Er führte darin für die bisher hier betrachtete KI das Akronym GOFAI (Good Old Fashioned Artificial Intelligence) ein und grenzte diesen bisherigen Mainstream der KI-Forschung von anderen Ansätzen ab, intelligente Maschinen zu konstruieren, die seit den 1980er Jahren immer größere Beachtung fanden. Dieses neue Forschungsprogramm wurde Konnektionismus genannt. Dabei geht es um viele gleiche oder ähnliche Elemente, die miteinander vernetzt sind und über Signale Nachrichten austauschen. Solche Modelle zur Simulation von intelligentem Verhalten sind beispielsweise Künstliche Neuronale Netze. Sie heißen so, weil sie den Neuronenverbänden in organischen Nervensystemen grob nachgebildet wurden. Die Grundidee stammt von Warren McCulloch und Walter Pitts, die schon 1943 ein logisches Modell von Nervenzellen und deren Vernetzung vorgestellt hatten, für



PD Dr. Rudolf Seising, Wissenschaftlicher Mitarbeiter im Forschungsinstitut für Technik- und Wissenschaftsgeschichte am Deutschen Museum, München

Alan Turing war davon überzeugt, dass zukünftige Computer das Leben der Menschen verändern würden und man um das Jahr 2000 wie selbstverständlich von „denkenden Maschinen“ sprechen würde, weil sich der Begriff „Denken“ gewandelt haben würde.

das sich bald darauf Nachrichtentechniker und Computerkonstrukteure interessierten, weil der bekannte Mathematiker John von Neumann die Terminologie dieses Artikels im Jahre 1945 nutzte, um die Konstruktion und Funktionsweise des neu zu bauenden Computers EDVAC zu beschreiben. Anstatt „Schalter“ schrieb er „neuron“ (Nervenzelle) und anstatt „Speicher“ schrieb er „memory“ (Gedächtnis). Einen Nachweis für eine prinzipielle Analogie von Gehirn und Computer gab es nicht, und es war klar, dass von Neumann hier stark vereinfacht hatte, doch ihn beschäftigte diese Thematik bis zum Ende seines Lebens, wie sein posthum unvollendet veröffentlichtes Buch *The Computer and the Brain* belegt.

Der Psychologe Frank Rosenblatt griff in den späten 1950er Jahren einen Vorschlag von McCulloch und Pitts auf, die von ihnen modellierten künstlichen Neuronen in die Lage zu versetzen, räumliche Muster zu erkennen, und er entwickelte das *Perceptron*, ein künstliches Neuronennetz, das zunächst Quadrate und Kreise, später Ziffern und Buchstaben „erkannte“. Zu Beginn der 1960er Jahre baute er mit seinen Mitarbeitern auch einen ersten realen Apparat namens *Perceptron Mark I* der mit seinem 20 x 20 Pixel großem Bildsensor Ziffern identifizierte, deren Merkmale er zuvor abgespeichert hatte.

Bereits Rosenblatt sprach bei dieser Mustererkennung aufgrund vieler Versuch-und-Irrtum-Verarbeitungen von einem „Lernalgorithmus“. Insofern war dies ein Anfang

der *New fashioned AI*, *New Wave AI* oder auch *new-fangled AI* (NFAI) genannten Forschung, die ihre Umwelt „wahrnehmen“, darauf reagieren, sich anpassen und „lernen“ konnte. Es sind diese NFAI-Systeme, die unseren heutigen Begriff der KI prägen.

Spiele Lernen

NFAI löste GOFAI nicht plötzlich ab, vielmehr überlagerten sich die Entwicklungen beider KI-Zugänge und sie werden auch

beide weiterhin verfolgt. Interessant ist zu sehen, wie weit sie tragen, wenn es darum ging, Computer Spiele mit festen Regeln spielen zu lassen, wobei die Güte des Spielens als Indikator für Intelligenz angenommen wird.

Schon in den 1940er Jahren dachte der deutsche Bauingenieur und Computerpionier Konrad Zuse darüber nach, wie er eine Rechenanlage nicht nur rechnen, sondern auch andere Probleme lösen lassen könnte. Dazu entwickelte er den Plankalkül, eine erste logische Sprache, um dem Rechner Befehle zu erteilen. Der Plankalkül sollte so allgemein sein, dass man damit auch die Regeln für das Schachspiel formulieren konnte. Zuses Vorarbeiten dazu finden sich heute im Konrad Zuse Internet Archive.

Um 1946 entwarf der britische Mathematiker Alan Turing gemeinsam mit seinem Studienfreund David Gawen Champernowne einen ersten Schach-Algorithmus, der *Turochamp* genannt und 1952 fertiggestellt wurde. Weil aber noch keine entsprechende Maschine zur Verfügung stand, konnte Turing ihn nicht als Programm

auf einem Computer laufen lassen. Er übernahm daher selbst die Rolle der Maschine, indem er ganz genau einen Befehl nach dem anderen ausführte, so wie sie zuvor auf Papier niedergeschrieben worden waren. Dabei spielte er gegen seinen Kollegen Alick Edwards Glennie. Turing brauchte damals zur Berechnung jedes Zugs etwa eine halbe Stunde!

Gemeinsam mit der Mathematikerin Cecily Popplewell (1920–1995) hatte Turing seit 1949 in seiner Abteilung für Computer Machine Learning an der University of Manchester Seminare durchgeführt. Der aus Deutschland geflohene und 1947 britischer Staatsbürger gewordene Dietrich Günther Prinz, der mit Turings Team eng zusammenarbeitete, hatte gelernt, den Rechner *Ferranti Mark 1* zu programmieren und er schrieb im November 1951 wohl das erste Schachprogramm *Matt in zwei Zügen*, das auf dieser Maschine lief und die Aufgabe in 15 Minuten löste. Prinz hatte sich von dem Artikel *A Theory of Chess and Noughts and Crosses* in den *Science News* im Jahre 1950 inspirieren lassen, den der britische Physiker und seit 1947 ebenfalls zu Turings Gruppe gehörige Donald Watts Davies geschrieben hatte. Bei jeder Spielpartie durchsuchte Prinz' Programm alle möglichen Züge und Gegenzüge, es wurden also tausende mögliche Spielzüge ausgewertet. Daher ließ das Computerprogramm mit fast 15 Minuten sehr viel länger auf seinen nächsten Zug warten, als ein versierter menschlicher Schachspieler. Der *Mark 1* war allerdings nicht mächtig genug, eine gesamte Schach-Partie zu spielen, deshalb hatte Prinz sein Programm auf die Spielsituation zwei Züge vom Matt entfernt begrenzt.

In den USA trug der US-amerikanische Bell-Mathematiker Claude Shannon am 9. März 1949 in New York über Ideen vor, Computer so zu programmieren, dass sie Schach spielen. Die schriftliche Fassung dieses Vortrags wurde im darauffolgenden Jahr in einer philosophischen Zeitschrift abgedruckt. Fast ein Jahrzehnt später animierten den IBM-Ingenieur Alex Bernstein Diskussionen mit Shannon dazu, den IBM-Rechner 704 Schach spielen zu lassen. Mit seinen Kollegen und unterstützt vom Schachgroßmeister Arthur Bernard Bisguier schrieb er ein Schachprogramm, das beim Dartmouth Treffen diskutiert wurde, obwohl es noch nicht fertiggestellt war. Im darauffolgenden Jahr lief dieses *Bernstein Chess Program* dann erfolgreich.

Dietrich Günther Prinz hatte gelernt, den Rechner *Ferranti Mark 1* zu programmieren und er schrieb im November 1951 wohl das erste Schachprogramm *Matt in zwei Zügen*, das auf dieser Maschine lief und die Aufgabe in 15 Minuten löste.

Modelle zur Simulation von intelligentem Verhalten sind beispielsweise Künstliche Neuronale Netze. Sie heißen so, weil sie den Neuronenverbänden in organischen Nervensystemen grob nachgebildet wurden. Die Grundidee stammt von Warren McCulloch und Walter Pitts.

Das Schachspiel wurde ein wichtiges Test- und Anwendungsfeld für KI-Programme. Schach ist ungeheuer komplex – „ultracompllicated“, wie Simon und Newell sich ausdrückten. Es war „the intellectual game par excellence“. Als die beiden mit dem Programmierer John Clifford Shaw 1958 die damals bekannten Schach-Programme untersuchten, kamen sie noch zu dem Urteil: „we have at least entered the arena of human play – we can beat a beginner.“ Schon bald darauf nannte der russische AI-Forscher Alexander Semenovich Konrod das Schachspiel die „Drosophila der AI“, eine Bezeichnung, die der Wissenschaftshistoriker und -soziologe Nathan Ensmenger vor etwa 12 Jahren gründlich diskutierte.

Spiel mit Dame

Das Spiel Schach führte die KI von der GOFAI zur NFAI. Bevor dieser Pfad nachgezeichnet wird, muss aber das Spiel Dame genannt werden, das diesen Wechsel früher einläutete, denn schon 1959 sah der Elektroingenieur Arthur Lee Samuel eine Möglichkeit, Computer „lernen“ zu lassen, ohne dass sie dafür explizit programmiert würden. Er demonstrierte dies mit seinem Programm, das die Regeln des Damespiels enthielt und sein Spiel jeweils verbessern konnte, weil erfolgreiche Spielzüge gegenüber nicht zielführenden Spielzügen höher bewertet wurden. Die mit den Spielzügen abgespeicherten Werte gaben die Wahrscheinlichkeit (lateinisch *probabilitas*) an, mit denen die Spielzüge gewinnbringend wären. Die entsprechenden Algorithmen heißen probabilistisch, da sie – anders als die klassischen (deterministischen) Algorithmen, die stets die eindeutige Lösung des Problems finden – nur wahrscheinliche Ergebnisse liefern. Samuel konnte so die Spielgüte seines Programms optimieren, obwohl das Programm selbst unverändert blieb.

Algorithmen, die nicht nach jedem Durchlauf denselben, sondern mit einer gewissen Wahrscheinlichkeit einen besseren Ausgabewert liefern, in diesem Sinne „lernen“, sind Werkzeuge des *Maschinellen Lernens* (ML) und damit der *New fashioned Artificial Intelligence* (NFAI).

Statistik

In den modernen Wissenschaften werden Wahrscheinlichkeiten zur Voraussage von Ereignissen gebraucht. Diese stützen sich auf genügend große Datenmengen, wie sie etwa zur Staatsbeschreibung, der ursprünglichen Bedeutung des Wortes „Statistik“ gesammelt und analysiert wurden.

In der Wissenschaftsdisziplin dieses Namens werden aufgrund verfügbaren Datenmaterials Hypothesen generiert und ihre Methoden sind inzwischen in den empirisch arbeitenden Wissenschaften unverzichtbar geworden. Die Geschichte des Fachs Statistik mündet zwischen mathematischer Theorie, wissenschaftlichem Berechnen und ihren verschiedenen Anwendungsgebieten, wie z. B. Geburts- und Sterberegister, Rentenversicherungen und Wirtschaftsdaten.

Mit letzteren begann ihre historische Entwicklung, denn zunächst wurden Daten der Anwendungsbereiche gesammelt und präsentiert. Das schlussfolgernde Vorgehen der Statistik wurde bald einer Mathematisierung unterzogen und dies wurde zum Fundament des mathematisch-statistischen Methodengebäudes bis zur Mitte des 20. Jahrhunderts. Bis dahin spielte der empirisch-numerische Zugang nur eine kleine Rolle. Dies änderte sich mit Aufkommen der elektronischen Computer, denn nun setzte ein Prozess ein, der die Statistik aus ihrem „Eigenbrötlerium“ um mathematische Strukturen herausgelöst und mitgerissen habe, argumentieren die Stanford-Statistik-Professoren Bradley Efron und Trevor Hastie in ihrem Buch *Computer Age Statistical Inference*.

Algorithmen, die nicht nach jedem Durchlauf denselben, sondern mit einer gewissen Wahrscheinlichkeit einen besseren Ausgabewert liefern, in diesem Sinne „lernen“, sind Werkzeuge des *Maschinellen Lernens* (ML) und damit der *New fashioned Artificial Intelligence* (NFAI).



Der Arbeitskreis bot den Teilnehmenden die Möglichkeit, in kleinerem Kreis mit dem Referenten Dr. Seising zu diskutieren.

Datenwissenschaft

Wie umfassend diese Veränderung der Statistik durch die Computer sein würde, war zur Mitte des 20. Jahrhunderts nicht abzusehen. Weitsichtig war allerdings schon im Jahre 1962 der Statistiker John Wilder Tukey, der in seinem Artikel *The Future of Data Analysis* forderte, die Statistik als Disziplin anwendungs- und berechnungsorientiert auszurichten. Zwar könne die Datenanalyse bei kleinen Datensätzen von Hand durchgeführt werden, aber die Geschwindigkeit und die Wirtschaftlichkeit der Antwort machten den Computer bei großen Datensätzen unentbehrlich und auch bei kleinen Datensätzen sehr

wertvoll. Darin bestehe die Zukunft der Datenanalyse, die sich in allen Bereichen der Wissenschaft und der Technologie als erfolgreich erweisen werde. 1982 plädierte Tukey in seinem Vortrag *Another Look at the Future* noch einmal ausdrücklich für eine intensive interdisziplinäre Zusammenarbeit der Kulturen der Statistik und der Computer Science.

Zwei Kulturen des Modellierens

In den 1970er Jahren suchte die *Environmental Protection Agency* (EPA) nach Methoden zur genaueren Voraussage der Ozonkonzentration in der Luft über dem Los-Angeles-Becken, die in den zehn Jahren vorher oft zu für die Bevölkerung gesundheitsgefährdenden Werten angestiegen war. Ziel der dazu in Auftrag gegebenen Studie war eine möglichst exakte Vorhersage der Ozonwerte für die jeweils nächsten 12 Stunden.

An Datenmaterial standen die über 450 meteorologischen Variablen zur Verfügung, die in den vergangenen sieben Jahren täglich und stündlich gemessen worden waren, sowie die zum jeweiligen Zeitpunkt gemessenen Ozonwerte. Aufgabe des Projekts war es, die mathematische Funktion zu schätzen, die den Ozonwert an einem Tag des Untersuchungszeitraums als Funktion der 450 Messwerte angibt. Dazu teilten die Projektarbeiter die Daten in zwei Datensätze auf: einen für das Training des Programms, der

aus den Daten der ersten 5 Jahre bestand, und mit dem ein Modell aufgebaut wird, und einen mit den Daten der letzten 2 Jahre, gegen den das Modell dann getestet wurde. Da mit den herkömmlichen statistischen Verfahren keine akzeptablen Ergebnisse erzielt werden konnten, scheiterte das Projekt.

Ein Mitarbeiter bei diesem Ozon-Projekt war der Statistiker Leo Breiman. Er schrieb aufgrund seiner hier gesammelten und weiterer Erfahrungen im Jahre 2001 den Artikel *Statistical Modeling: The Two Cultures*, in dem er konstatierte, dass die Statistiker*innen

mittlerweile zwei unterschiedliche „Kulturen“ hervorgebracht hatten: 1) die Kultur des statistischen Modellierens, bei der sie ausgehend von der Prämisse, dass ein gegebenes stochastisches Datenmodell die Daten erzeuge, von den Daten zu ihren Schlussfolgerungen gelangten; 2) die Kultur des algorithmischen Modellierens, bei der ein Algo-



Die Referierenden Dr. Anna Puzio, University of Twente (Niederlande), und Prof. Dr. Alexander Filipović, Professor für Christliche Sozialethik an der Universität Wien, beantworteten im Anschluss an Ihren Vortrag Fragen aus dem Plenum.

rithmus ohne Vorannahmen über einen etwaigen Datenmechanismus die Testdaten klassifiziert und aufgrund der schon vorliegender Daten zu Voraussagen kommt. Diese Kultur des algorithmischen Modellierens hatte sich vor allem außerhalb der Universitäten als alternatives Vorgehen bei hochkomplexen Problemen entwickelt und etabliert.

Breiman hielt die erste Kultur für die zu lösenden Probleme ungenügend, weil die betrachteten Systeme ungeheuer groß und komplex sind und aus den Wissenschaften immer mehr Fragen aufkamen, die nicht beantwortet werden konnten. So mussten auch die Datenstrukturen immer komplexer werden und es wurde schwieriger, geeignete Datenmodelle zu konstruieren. Die zweite Kultur konnte hingegen wegen ihrer hohen Prognosegenauigkeit immer mehr Erfolge vorweisen.

Den von Breiman geforderten Schwenk der Statistik weg von der Wahrscheinlichkeitstheorie und hin zu Algorithmen nannte der mit ihm befreundete Stanford-Professor Jerome Herold Friedman eine „Data-Mining-Revolution“. Die Statistik sei an einem Scheideweg und er empfahl seiner Zunft: „make peace with computing“ und „moderate our romance with mathematics“.

Im gleichen Jahr forderte Chien-Fu Jeff Wu, Statistik-Professor in Michigan, seine Community auf, das Fach Statistik in Data Science umzubenennen und nicht mehr von „Statistiker*innen“, sondern von „Data scientists“ zu sprechen: „It is time in the history of statistics to make a bold move“: Man möge sich auf die „großen Datenmengen“ fokussieren, sich den anderen Wissenschaften mehr öffnen – auch für die Ausbildung von Datenwissenschaftler*innen sollten die anderen Wissenschaften treibend sein –, deren empirisch-physikalischen Ansatz und deren Wissen zur Problemlösung nutzen.

„Something important changed in the world of statistics in the new millennium“, schrieben Efron und Hastie. Die Ursache dafür sahen sie in den Möglichkeiten der Voraussagealgorithmen, die das Fach Statistik zu Data Analytics und schließlich zur Data Science wandelten. Mit ihren Erfolgen, die sich bei enorm großen Datenmengen einstellten, wurden die Algorithmen immer wichtiger. Efron und Hastie nennen sie die „media stars of the Big-Data era“.

Leo Breiman konstatierte, dass die Statistiker*innen mittlerweile zwei unterschiedliche „Kulturen“ hervorgebracht hatten: 1) die Kultur des statistischen Modellierens; 2) die Kultur des algorithmischen Modellierens. Die zweite Kultur konnte wegen ihrer hohen Prognosegenauigkeit immer mehr Erfolge vorweisen.

GOFAI und Schach

Garri Kimowitsch Kasparow war von 1985 bis 2000 Schachweltmeister, der auch oft Wettkämpfe mit Turnierbedenkzeit gegen Schachprogramme bestritt. Gegen den von der Firma IBM gebauten *Deep Thought* spielte er 1989 zwei Partien erfolgreich und auch dessen Nachfolger *Deep Blue* besiegte er im Jahre 1996 mit 4:2. Im Jahr darauf verlor er allerdings den Rückkampf mit 2,5:3,5.

Der 1985 in die USA eingewanderte Taiwanese Feng-hsiung Hsu, der 1989 mit einer Arbeit über Computerschach an der Carnegie Mellon University in Pittsburgh (Pennsylvania) promoviert wurde, konstruierte darauf aufbauend mit seinem Team den Computer *Deep Thought*. Dessen Nachfolger *Deep Blue* entstand dann bei IBM in Zusammenarbeit von Hsu mit dem kanadischen Informatiker Murray S. Campbell. *Deep Blue* enthielt riesige Schach-Datenbanken, hatte enorme Rechenleistung und konnte enorm viele Züge und Kombinationen simulieren. Diese große Rechenkapazität, riesige Mengen von möglichen Zügen im Voraus durchzuprobieren, gelang sozusagen mit „Brute Force“ (roher Gewalt).

NFAI und Go

Weitaus komplexer als Schach und damit die weitaus größere Herausforderung für die Entwickler von KI-Systemen war das japanische Spiel Go, dessen Brett mit seinen 19 x 19 Feldern größer als das 8 x 8 Feld des Schachspiels ist und ungleich mehr mögliche Züge bietet. Bei diesem Spiel versuchen beide Spieler mit ihren Steinen mehr Gebiete auf dem Spielfeld zu erobern als der Gegner. Dazu müssen die Gebiete mit eigenen Steinen umzingelt werden. Zudem bringt das Schlagen von gegnerischen Steinen Punkte.

Mit traditionellen Brute-Force-Algorithmen war das Go-Spiel gegen gute Spieler nicht zu gewinnen. Im Gegensatz zu Schach gab es für Go aber auch keine zweckmäßigen heuristischen Methoden, um eine gegebene Spielstellung zu bewerten. Erst das algorithmische Modellieren führte hier zum Ziel, als das Computerprogramm *AlphaGo* im Oktober 2015 den mehrfachen Go-Europameister Fan Hui (2. Dan) unter Turnierbedingungen besiegte. *AlphaGo* von der britischen Firma DeepMind nutzte ein künstliches neuronales Netzwerk für „Deep Learning“. Es kam damit einem von der Firma Facebook entwickelten System um wenige Wochen zuvor, denn am 27. Januar 2016 hatte Mark Zuckerberg gepostet, dass man dort einem Computer das Go-Spiel beigebracht habe. Noch im gleichen Jahr schlug *AlphaGo* den Südkoreaner Lee Sedol, einen der weltbesten Go-Profispieler,

ChatGPT wurde mit großen Datenmengen trainiert, und wenn wir dieser neuen KI (NFAI) nun Fragen oder Anweisungen übermitteln, so verarbeitet und analysiert es diese unter Nutzung seiner Trainingsdaten in seiner neuronalen Netzwerk-Architektur.

er, nachdem es zuvor mit 30 Millionen Go-Partien trainiert wurde, die menschliche Meister gespielt hatten.

Im Jahr 2017 stellte ein Team um Demis Hassabis und David Silver von Googles KI-Forschungszentrum DeepMind in London im Wissenschaftsmagazin *Nature* ein Nachfolgeprogramm vor: *AlphaGo Zero*. Dieses Programm brauchte keinen menschlichen Lehrer mehr. Es erhielt die Go-Regeln und „lernte“, indem es gegen sich selbst spielte. Ein neuronales Netzwerk und Voraussagealgorithmen bestimmten den jeweils nächsten Zug. Gute Züge wurden belohnt – man spricht von verstärkendem maschinellem Lernen.

Auch NFAI denkt nicht

Auch ChatGPT – die Abkürzung für Generative Pre-trained Transformer (ein Umwandler, der zuvor trainiert wurde und Texte erzeugt) – beruht auf dem algorithmischen Modellieren



Eine noch größere Herausforderung für Entwickler von KI-Systemen war das japanische Go-Spiel. Die deutlich höhere Anzahl an Spielfeldern als beim Schach bietet weitaus mehr mögliche Züge und macht die Berechnung so komplexer.

des maschinellen Lernens. ChatGPT wurde mit großen Datenmengen trainiert, und wenn wir dieser neuen KI (NFAI) nun Fragen oder Anweisungen übermitteln, so verarbeitet und analysiert es diese unter Nutzung seiner Trainingsdaten in seiner neuronalen Netzwerk-Architektur. Die relevanten Informationen werden herausgestellt und eine Antwort erzeugt, die das KI-System an uns zurücksendet. ChatGPT antwortet also aufgrund eines umfangreichen Trainings mit großen Datenmengen, darunter sind ein Textkorpus aus Büchern, Briefen, Wikipedia-Einträgen oder auch literarischen Textsammlungen, sowie das gesamte Gutenberg-Projekt.

Die Antworten wurden nicht von ChatGPT erdacht, sondern sie sind als Wortfolgen entstanden, die nach algorithmisch gewonnenen Wahrscheinlichkeiten auf diese Weise zusammengesetzt wurden. Einen Sinn in diesen Antworten gibt es erst dann, wenn *wir* ihn darin erkennen, also wenn *wir* ihn denken, denn Künstliche Intelligenz kann nicht denken! ■

Sprachmodelle, Wahrhaftigkeit und die soziale Natur des Wissens

Künstliche Intelligenz in einer demokratischen Gesellschaft
von Karoline Reinhardt

Wie sollte KI-Entwicklung in einer demokratischen Gesellschaft ausgestaltet sein? Welche Auswirkungen haben gegenwärtige KI-Entwicklungen auf Demokratie und demokratische Prozesse? Um diese Fragen auszuloten, werde ich mich im Folgenden auf ein jüngeres und viel diskutiertes Beispiel aus der aktuellen KI-Forschung beziehen: nämlich generative Sprachmodelle, die auf *Large Language Models (LLMs)* beruhen.

Einleitung

Eines der bekanntesten davon ist ChatGPT.

Bevor ich aber zur philosophisch-ethischen Einordnung dieser Sprachmodelle komme, gilt es zunächst herauszuarbeiten, inwiefern philosophische Ethik eigentlich Einschätzungen zu technischen Artefakten liefern kann. Was kann eine ethische Perspektive auf Technologien und insbesondere auf computerisierte, probabilistische Modelle leisten, wie sie uns in Form von Large Language Models und den auf ihnen basierenden Anwendungen ChatGPT und Bard begegnen? Verfügt die philosophische Ethik überhaupt über ein Instrumen-



Prof. Dr. Karoline Reinhardt,
Juniorprofessorin für Angewandte Ethik
an der Universität Passau

tarium, mit dem wir diese Modelle und Anwendungen evaluieren, gegebenenfalls kritisieren können – und sogar Aussagen dazu treffen können, wie diese gestaltet werden *sollen*? Und wann und wie sie eingesetzt werden sollen – und wann nicht? Oder können wir ihre Funktionsweise, ihre Nutzung, ihre Wirkung lediglich beschreiben? Lassen sich Sprachmodelle überhaupt sinnvoll ethisch bewerten?

Einführend werde ich auf Large Language Models als Unterform probabilistischer Modelle, die auf dem Maschinellen Lernen beruhen, eingehen. Anschließend werde ich einige Herausforderungen herausarbeiten, die mit diesen verbunden sind, und welche Perspektiven Technikethik ermöglicht, um diese Herausforderungen zu beschreiben und zu bearbeiten.

Large Language Models als probabilistische Modelle

Im Zentrum der Aufmerksamkeit der Software-Entwicklungen, die unter der

Überschrift „Künstliche Intelligenz“ (KI) gefasst werden, stehen aktuell Algorithmen, die neue Bilder oder Texte generieren (wie etwa ChatGPT von OpenAI oder Bard von Google) – und zwar gepromptet (von engl. *to prompt* veranlassen) durch Eingaben von Nutzer:innen in natürlicher Sprache. Die dabei generierten Texte sind nicht einfach kopiert, sondern werden tatsächlich von diesen Anwendungen in der Interaktion erzeugt.

Chatbots wie ChatGPT beruhen auf Large Language Models. Diese Modelle wurden auf der Grundlage von Milliarden von (zumeist) englischsprachigen Dokumenten entwickelt. In aufwendigen Auswertungs- und Trainingsprozessen wurden mithilfe von Maschinellen Lernen in diesen Dokumenten Muster aufgefunden und erfasst. Diese Muster werden dann über die Chatbots reproduziert. Die Bausteine für die Mustererkennung und -reproduktion sind dabei Buchstabensequenzen (*tokens*), die in Zahlenwerte transformiert werden. Anschließend wird berechnet, wie sich die Position verschiedener Buchstabensequenzen im Kontext der anderen verhält. Aufgrund der Nähe und Entfernungen zwischen den einzelnen Buchstabensequenzen wird auf diese Weise errechnet, mit welcher Wahrscheinlichkeit eine bestimmte Sequenz auf die nächste folgt.

Die mit diesen Modellen arbeitenden Chatbots verfügen, das dürfte mittlerweile hinreichend bekannt sein, nicht über ein menschliches Sprachverstehen: Sie verstehen weder die Bedeutung dessen, was man als Nutzer:in eingibt, noch was sie anschließend als Text produzieren. Die Texterstellung beruht also nicht auf Textverständnis, sondern auf Mustererkennung und Musterreproduktion.

Die ChatGPT zugrunde liegende Technologie ist nicht grundsätz-

Chatbots wie ChatGPT beruhen auf Large Language Models. Diese Modelle wurden auf der Grundlage von Milliarden von (zumeist) englischsprachigen Dokumenten entwickelt. Mithilfe von Maschinellen Lernen wurden in diesen Dokumenten Muster aufgefunden und erfasst.

lich neu. Sie beruht auf den gleichen Grundlagen wie alle Anwendungen, die so Maschinelles Lernen zum Einsatz bringen: sich optimierende Neuronale Netze, die auf der Grundlage einer Vielzahl sich im Trainingsprozess justierender Parameter interne Modelle bilden. Neu jedoch sind die riesigen Datensätze, die beim Training eingesetzt wurden und werden, und die es letztlich ermöglichen, natürlichsprachlich mit der Anwendung zu interagieren und Texte zu produzieren, die (kaum mehr) von Texten zu unterscheiden sind, die Menschen verfasst haben.

Möglich ist dies aufgrund der Auswertung sehr großer Datenkorpora mit Hilfe statistischer Methoden. Während die klassischen Problemlösungsalgorithmen, die etwa vom schriftlichen Rechnen bekannt sind, bei korrekter Anwendung *notwendig* zur Lösung führen, liefern prädiktive, statistisch arbeitende Algorithmen *wahrscheinliche* Outputs: Das Ergebnis kann stimmen – oder auch nicht. Es hat eine hohe Wahrscheinlichkeit, ist aber nicht wahr.

Wichtig ist darüber hinaus noch, dass alle von ChatGPT erzeugten Texte weder auf Faktenwissen noch auf Expertise beruhen. ChatGPT weist hierauf zum Beispiel auch an verschiedenen Punkten hin. So liest man zu Beginn der Nutzung von ChatGPT, dass

obwohl bestimmte Schutzmaßnahmen implementiert seien, das System gelegentlich falsche oder irreführende Informationen liefern und beleidigende oder parteiische Inhalte produzieren kann. Es sei nicht dazu gedacht, Ratschläge zu erteilen.

Sozialität des Wissens

Damit bringt uns ChatGPT an einen Grenzpunkt menschlicher Wissensgenese und -akkumulation: Der überwiegende Teil unserer individuellen-menschlichen Wissensbestände basiert auf den Erkenntnissen und Aussagen anderer. Fast alles, was wir wissen, wissen wir aufgrund der Aussagen oder textlichen und bildlichen Erzeugnisse, die andere uns mitteilen oder hinterlassen haben. In nur sehr eingeschränkten Bereichen können wir gesichertes Wissen erlangen, ohne auf das Wissen und die Erkenntnisse anderer zurückzugreifen. Und dieser Umstand wird umso relevanter, je ausdifferenzierter menschliche Gesellschaften sind: Haben Sie heute Morgen Zeitung gelesen? Sind Sie Auto gefahren? Haben Sie vielleicht sogar Ihre Kinder zum Kindergarten oder in die Schule gebracht? Stellen Sie sich vor, Sie hätten alle Daten und Informationen selbst erheben, die Technologien in ihrer Zuverlässigkeit selbst prüfen, gegebenenfalls Ihre Kinder

Fast alles, was wir wissen, wissen wir aufgrund der Aussagen oder textlichen Erzeugnisse, die andere uns mitteilen oder hinterlassen haben. In nur sehr eingeschränkten Bereichen können wir gesichertes Wissen erlangen, ohne auf das Wissen anderer zurückzugreifen.

nicht in die Betreuung anderer übergeben und alles Wissen, was diese in der Schule erlernen, sich erst einmal selbst erarbeiten müssen. Wenn man sich überlegt, was man im Laufe eines Tages an Informationen genutzt hat, um den Tag zu planen und ihn mehr oder erfolgreich zu gestalten, und man sich anschließend überlegt, man hätte all diese Informationen und die ihnen zugrundeliegenden Daten erstmal selbst erheben, prüfen, einordnen und interpretieren müssen, um letztlich zu Schlüssen und Entscheidungen zu gelangen, dann bekommt man recht schnell ein Gefühl dafür, wie wichtig Vertrauen für ausdifferenzierte soziale Gefüge ist.

Das Vertrauen in andere und Institutionen, zum Beispiel Forschungseinrichtungen, Universitäten, Verlage, ermöglicht das Voranschreiten in der Wissensproduktion, das wir erleben, und die gesellschaftliche Ausdifferenzierung, die uns in unserer Lebenswelt vielleicht allzu selbstverständlich scheint. Gleichzeitig macht diese Ausdifferenzierung auch ein hohes Maß an Vertrauen nötig, um in unserer komplexen Lebenswelt handlungsfähig zu bleiben.

Das ist auch einer der Gründe, warum Wahrhaftigkeit und Vertrauenswürdigkeit einen so hohen gesellschaftlichen Stellenwert haben, warum Lüge und Vertrauensbruch auch moralisch so stark geahndet werden, warum uns Fake News und strategische Falschaussagen so beunruhigen: Wahrhaftigkeit bzw. deren Fehlen führt uns an den Rand der menschlichen Lebensweise, so wie wir sie hier und heute kennen. Und hier deutet sich nun schon an, warum all dies für Demokratien äußerst relevant ist.

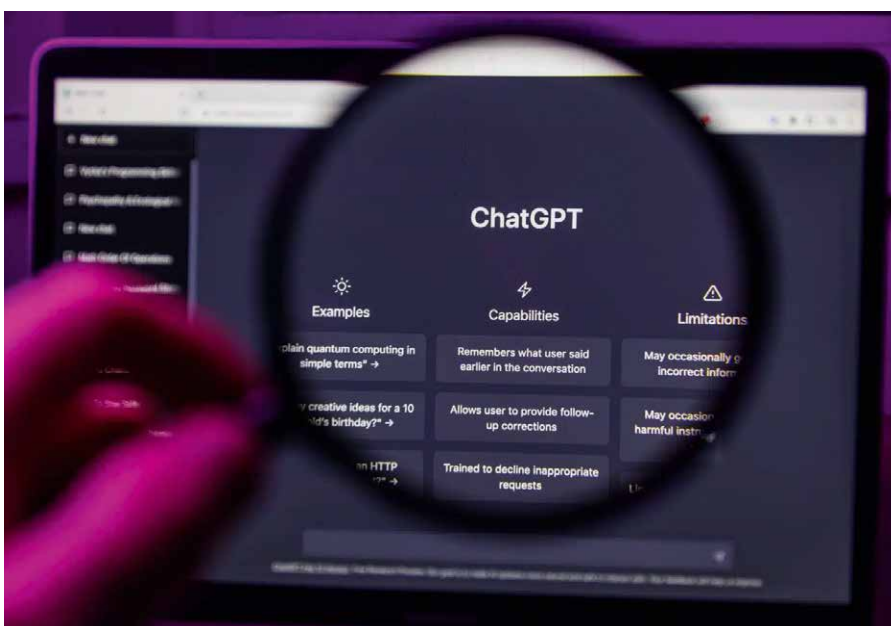


Foto: Marketcomlabo / Wikimedia Commons, CC BY-SA 4.0

Alle von ChatGPT erzeugten Texte beruhen weder auf Faktenwissen noch auf Expertise, sondern auf Wahrscheinlichkeiten. Das System liefert gelegentlich falsche oder irreführende Informationen, sodass immer eine kritische Prüfung durch den Nutzer nötig ist.

Wahrhaftigkeit, Kommunikation und Demokratie

Öffentliche Kommunikation ist ein Kernelement von Demokratie, die Bestimmung des Gemeinwohls wird aus öffentlichem Diskurs legitimiert. Dieses Verständnis von Öffentlichkeit und deren Funktion geht mit einer Reihe von normativen Erwartungen und Annahmen einher: etwa mit der Erwartung, dass die geteilte Zielsetzung die Aushandlung und Verständigung ist; mit der Annahme der Achtung der Selbstzweckhaftigkeit aller Beteiligten und der wechselseitigen Anerkennung der Gleichberechtigung der Kommunizierenden in der Kommunikationssituation als Gelingensbedingung von Verständigung; aber eben auch Erwartungen von Wahrheit und Wahrhaftigkeit.

ChatGPT und andere Large Language Models stellen für dieses Verständnis von Öffentlichkeit eine

Kommunikationssituation? Wer sind die Autor:innen welcher Inhalte? Bei welchen Kommunikationsakten handelt es sich um die Äußerungen von Kommunikationsteilnehmer:innen, die sich den genannten Grundsätzen gelungener Kommunikation verpflichtet sehen? Eventuell könnte dieses Misstrauen auch zu einer generellen Zurückhaltung in Bezug auf die aktive und passive Teilnahme an öffentlicher Kommunikation führen (Heesen 2023, ebd.). Gleichzeitig können Anwendungen wie ChatGPT auch bestimmte Gruppen befähigen überhaupt erst am öffentlichen Diskurs teilzunehmen und in diesem Gehör zu finden.

Selbst wenn man diese Überlegungen teilt, könnte man dann nicht immer noch argumentieren, dass ChatGPT halt „nur ein Werkzeug“ sei, und dass es eben darauf ankomme, dass wir lernen, richtig damit umzugehen. Verfügt Ethik beziehungsweise Technikethik überhaupt über ein In-

ChatGPT stellt für unser Verständnis von Öffentlichkeit eine Herausforderung dar: Die Kommunikationssituation, in der wir uns in der öffentlichen Kommunikation befinden, wird nun auch von generierten Inhalten mitgestaltet wird.

menschliches Handeln einer ethischen Bewertung unterzogen werden kann, oder erweitert vielleicht noch um institutionelles Handeln. Artefakte, wie Hammer, Zange, Bohrmaschine – oder eben ChatGPT – seien dieser Bewertung dagegen entzogen.

Der Einschätzung „Es ist nur ein Werkzeug“ möchte ich im Folgenden drei Grundüberzeugungen der Technikethik entgegensetzen, die ich zunächst erläutern und anschließend auf generative Sprachmodelle wie ChatGPT anwenden werde: (1) Technik ist nicht neutral. (2) Technik ist gerichtet. (3) Technik existiert nicht im Vakuum.

Die erste These möchte ich über ihr Gegenteil erläutern, nämlich die Neutralitätsthese, die besagt: Technik sei lediglich ein neutrales Mittel zur Erreichung bestimmter Zwecke. Sie kann freilich, so die Neutralitätsthese weiter, zu guten wie auch zu schlechten Zwecken eingesetzt werden. Diese Zwecke können einer moralischen Bewertung unterzogen werden, die Technik selbst bleibt aber einer moralischen Bewertung entzogen: Den Hammer kann ich, nach der Neutralitätsthese, benutzen, um einen Nagel in die Wand oder um jemandem den Schädel einzuschlagen. Der Hammer selbst verhielte sich aber neutral zu diesen Zwecken. Bezogen auf ChatGPT hieße dies: Ich kann diese Technologie nutzen, um massenhaft Falschmeldungen zu erstellen oder ich kann diese Technologie nutzen, um sprachlich eingeschränkte Personen in der Kommunikation z. B. mit Ämtern bei der Erstellung von Schriftstücken zu unterstützen. Nach der Neutralitätsthese lässt sich allein die Nutzung ethisch bewerten, nicht aber das Werkzeug/die Technologie selbst.

Dagegen geht die Technikethik davon aus, dass Technologie nicht neu-



Collage: NanoStockk / canva.com

Öffentliche Kommunikation ist ein Kernelement von Demokratie. Die Kommunikationssituation wird aber immer mehr von KI-generierten Inhalten mitgestaltet. Dies könnte zu Misstrauen gegenüber der Medienkommunikation führen. Autorinnen und Autoren müssen im Kommunikationsprozess erkennbar sein.

Herausforderung dar: In Massendemokratien wird öffentliche Kommunikation zumeist medial vermittelt organisiert und strukturiert. Da die Kommunikationssituation, in der wir uns in der öffentlichen Kommunikation befinden, die nun auch von generierten Inhalten mitgestaltet wird, unübersichtlicher wird, könnte dies zu Misstrauen gegenüber Medienkommunikation führen (Bernd Heesen, *Künstliche Intelligenz und Machine Learning mit R*, 2023, 12). Wer sind eigentlich die Akteur:innen in dieser

strumentarium, um Large Language Models und die auf ihnen basierenden Anwendungen ethisch zu bewerten?

Nur ein Werkzeug?

Inwiefern ist ChatGPT tatsächlich nur ein Werkzeug? Was heißt es eigentlich, dass etwas nur ein Werkzeug ist? Eine Weise, in der dieser Satz verwendet wird, impliziert, dass eine ethische Bewertung von Werkzeugen nicht angemessen sei. Dass letztlich immer nur Menschen handeln würden und allein

tral ist. Technologie verfüge vielmehr über intrinsische Eigenschaften, die bestimmte Nutzungsweisen ermöglichen und andere ausschließen: Mit einem Hammer kann ich einen Nagel in die Wand schlagen, mit einem Wattebausch ist das schwieriger. Umgekehrt ist der Hammer nicht geeignet, um damit eine Wunde mit Desinfektionsmittel abzutupfen.

Die Einschreibungen bestimmter Nutzungsmöglichkeiten können bewusst oder unbewusst, absichtlich oder unabsichtlich erfolgen. Aber sie sind da. Da in Technologie bestimmte Nutzungsmöglichkeiten eingeschrieben und andere ausgeschlossen werden, können wir nicht nur allein die Zwecke ethisch bewerten, zu denen sie eingesetzt werden, sondern auch die Technologie selbst. Jemand, der das recht früh im Hinblick auf Computer untersucht hat, war Joseph Weizenbaum. Weizenbaum war der Entwickler des Sprachprogramms ELIZA, das psychotherapeutische Gespräche nach der Methode von Carl R. Rogers simulieren konnte und schon in den 1970er Jahren Diskussionen ausgelöst hat, ob man nicht Psychotherapie an solche Programme auslagern könne. Ausgehend von der Beobachtung, dass Werkzeuge immer auch Welt strukturieren, stellt er sich in seinem Buch *Die Macht der Computer und die Ohnmacht der Vernunft* (1978) die Frage, welchen Einfluss Computer auf unsere Welt- und Selbstwahrnehmung haben. In anderer Weise betont Ruha Benjamin in ihrem Buch *Race after Technology* (2019), dass Technologie nicht neutral sei. Sie reflektiert vielmehr die sozialen und rechtlichen Codes, die in technischen Systemen eingebettet sind,

Da in Technologie bestimmte Nutzungsmöglichkeiten eingeschrieben und andere ausgeschlossen werden, können wir nicht nur allein die Zwecke ethisch bewerten, zu denen sie eingesetzt werden, sondern auch die Technologie selbst.



Prof. Dr. Michael Reder, Professor für Praktische Philosophie an der Hochschule für Philosophie München und Leiter der Philosophischen Tage, saß mit Prof. Dr. Karoline Reinhardt auf dem Podium. Gemeinsam beantworteten und diskutierten sie die Fragen der Teilnehmenden.

und betont, dass technologische Entwicklungen bestimmten Perspektiven und Formen der sozialen Organisation entspringen (Benjamin 2019, 77).

Die zweite These lautet, dass Technik immer gerichtet ist. Sie ist zweckbestimmt und zielgerichtet, da sie für die Lösung bestimmter Problemlagen geschaffen wird. Werkzeuge unterliegen einer „um-zu“-Logik, sie werden für eine bestimmte Aufgabe geschaffen, für einen bestimmten Zweck. Die Zwecke, für die Technologien geschaffen werden, ergeben sich aber nicht naturläufig, sie werden gesetzt. Zwecksetzungen können ethisch eingeordnet werden. Dort, wo sie zum Beispiel andere betreffen, ist ihre Überprüfung nach ethischen Gesichtspunkten sogar geboten. Die Idee, dass Verzerrungen durch Technologien unbeabsichtigt oder unbewusst sind, wird von Benjamin in Frage gestellt. Sie argumentiert, dass jede Technologie eine Intention habe, weil man nichts gezielt erschaffen könne, ohne eine bestimmte Absicht zu verfolgen, oder ohne sich jene vorzustellen, die diese Technologie letztlich zu irgendetwas nutzen sollen (ebd., 28). Selbst die Entscheidung, was denn als Problem behandelt wird, für das eine Lösung gefunden werden müsse, beruht auf vielfältigen Vorannahmen (ebd., 11), die nicht einfach neutral sind.

Die dritte These besagt, dass Technik nicht im luftleeren Raum entwickelt und verwendet wird. Sie ist vielmehr eingebettet in Produktions- und Nutzungsbedingungen sowie politische und rechtliche Rahmenbe-

dingungen. Technologie und Technologieentwicklung sind in menschliche Praktiken eingebunden, gleichzeitig prägen sie diese Praktiken und das menschliche Selbstbild wie auch Selbstverhältnis nachhaltig. Technologien werden dabei durch die Rahmenbedingungen beeinflusst, wie auch die Rahmenbedingungen durch die Technologien: Werte, Normen, Ideen werden in Technologien eingeschrieben. Gleichzeitig strukturieren Technologien auch Welt, erschaffen sogar neue Wirklichkeiten: „Der Computer“ ist dann zum „unentbehrlichen Bestandteil jeder Struktur“ geworden, heißt es bei Weizenbaum, „sobald er so total in die Struktur integriert ist, so eingesponnen in die verschiedensten lebenswichtigen Substrukturen, dass er nicht mehr herausgenommen werden kann, ohne unweigerlich die Gesamtstruktur zu schädigen“ (Weizenbaum 1978, 49f.).

An diesem Punkt befinden wir uns. Wir können hinter diesen Schritt nicht mehr zurückgehen, aber wir sollten diese Prozesse aktiv gestalten und nicht einer Logik der Machbarkeit unterliegen. Die Rahmenbedingungen von Technologieentwicklung müssen daher mit in den Blick genommen werden – und die Auswirkungen von Technologieentwicklung auf die Rahmenbedingungen ebenfalls.

ChatGPT: It's just a tool?!

Ist ChatGPT also nur ein Werkzeug? Ja, vielleicht schon. Vorausgesetzt,

Ist ChatGPT also nur ein Werkzeug? Ja, vielleicht schon. Vorausgesetzt, dass wir in den Blick nehmen, dass in Werkzeuge bestimmte Nutzungsweisen eingeschrieben sind, dass sie für bestimmte Zwecke gemacht werden.

dass wir in den Blick nehmen, dass in Werkzeuge bestimmte Nutzungsweisen eingeschrieben sind, dass sie für bestimmte Zwecke gemacht werden, von den Rahmenbedingungen ihrer Entstehung geprägt sind und profunde Auswirkungen auf diese haben können. Dass Werkzeuge also große normative Macht ausüben können. Wir müssen uns daher ganz genau anschauen, welche Wertsetzungen in diese Technologie eingeschrieben sind. Wir sind noch dabei zu verstehen, wie genau LLMs funktionieren, welche Fähigkeiten und welche Grenzen sie haben – und von welchen technischen Faktoren diese jeweils abhängen, aber einige Aspekte werden bereits jetzt diskutiert: LLMs tragen weiter dazu bei, dass die Verwendung der Daten Dritter zur unautorisierten Weiternutzung eine Normalität darstellt. Sie tragen außerdem zu einer Homogenisierung von Sprech- und Ausdrucksweisen bei und dazu, dass bestimmte Sprachen und Sprechweisen als Standard und andere als Abweichung betrachtet werden. Die Entwicklung und Verbreitung von LLMs, wie wir sie erlebt haben, setzt außerdem eine bestimmte Art von Technikentwicklung und -einsatz fort, die nicht demokratisch rückgebunden ist, auch wenn sie gegebenenfalls massive gesellschaftliche Folgen hat. (Damit ist freilich nicht gesagt, dass demokratische Rückbindung von Technikentwicklung nie zu problematischen Folgen führen würde.) Außerdem befördern die Anwendungen, die auf LLMs beruhen, eine weitere Individualisierung von Kommunikation und Ratsuche: Man wendet sich mit Fragen nicht an andere Menschen, sondern an die Chatbots. Einordnen muss man die Antworten in beiden Fällen. Auch Menschen irren sich, machen Falschaussagen, aber es fehlt das Moment der Intersubjektivität.

Damit geht unter anderem einher, dass wir in der Kommunikation mit einem Chatbot nicht mehr aufgefordert sind, uns Gedanken über andere zu machen; non-verbale Zeichen und Signale zu lesen.

Um auf die These von der Sozialität unseres Wissens zurückzukommen: Auch wenn Chatbots sich nicht in dem Umstand von Menschen unterscheiden, dass sie falsche Aussagen treffen – denn Menschen tun dies schließlich auch – besteht in der technischen Festschreibung ein entscheidender Unterschied: Normen von Wahrheitsfähigkeit und Wahrhaftigkeit werden hier technisch unterminiert, nicht als Fehler, Irrtum oder auch aus Böswilligkeit, so wie wir es je schon von Menschen kennen, sondern als Bestandteil der Technologie. Dass Wahrheit und Wahrhaftigkeit keine Rolle spielen ist Bestandteil der probabilistischen Struktur der Modelle. Dieser Umstand verletzt normative Kernannahmen für gelingende Kommunikation.

Plädoyer für Integrierte Technikentwicklung

Insgesamt zeigen diese Überlegungen zu generativen Sprachmodellen, dass Techniken nicht nur Werkzeuge sind, sondern auch gestaltender Faktor unserer Welt. Ihre Entwicklung und Anwendung bedürfen einer kritischen Reflexion im Hinblick auf ethische Aspekte, demokratische Werte und gesellschaftliche Auswirkungen. Gesellschaftliche Debatten über KI

und deren Anwendungen sind aber nur möglich, wenn bekannt ist, wie diese funktionieren, wo sie zum Einsatz kommen und welche Auswirkungen dies hat – und auch Zeit zur Auseinandersetzung mit den möglichen Auswirkungen ist, bevor sie zum Einsatz kommen.

Aktuell werden wir immer wieder mit neuen technologischen Entwicklungen konfrontiert, sobald diese sich auf dem Markt befinden. Ethische, soziale und rechtliche Überlegungen sollten aber nicht *ad hoc* und auch nicht erst *ex post* in die Betrachtungen mit einbezogen werden, sondern

Gesellschaftliche Debatten über KI und deren Anwendungen sind nur möglich, wenn bekannt ist, wie diese funktionieren, wo sie zum Einsatz kommen und welche Auswirkungen dies hat.

im Sinne einer integrierten Forschung und Technikentwicklung bereits in die kritische Betrachtung der Problemdefinition, die Ausgestaltung des Entwicklungsprozesses und die Implementierung. Weil Technik eben nicht neutral ist, weil die Zwecksetzungen in unserer Hand liegen, weil Technik nicht im luftleeren Raum entsteht und bleibt. ■



In der Studienbibliothek tagte der Arbeitskreis mit Professorin Karoline Reinhardt. Hier wurden Fragen und Gedanken zur KI ausführlich diskutiert.