

# Künstliche Intelligenz

## Überblick und gesellschaftlicher Ausblick

Frank Puppe

Lehrstuhl für Künstliche Intelligenz und Wissenssysteme der Universität Würzburg

Seit es Computer gibt, wird versucht, diese intelligent zu gestalten. Die erste und noch immer aktuelle Definition von Künstlicher Intelligenz (KI) orientiert sich am Menschen und stammt von Alan Turing aus dem Jahr 1950: „Ein Computer ist intelligent, wenn ein Mensch nach einer gewissen Zeitspanne in freier Unterhaltung mit ihm nicht unterscheiden kann, ob es ein Mensch oder ein Computer war.“ Es gibt auch menschenunabhängige Definitionen der KI, die zielgerichtetes Handeln betonen: „Any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals“<sup>1</sup>. Hier hängt der Grad der Intelligenz von der Komplexität der für die Zielerreichung relevanten Umgebung ab. Der Unterschied ist gravierend. Nach der ersten Definition könnte man den Computer als Gesprächspartner einfach entlarven, indem man ihn nach der Multiplikation zweier sechsstelliger Zahlen fragt. Wenn er das schnell und fehlerfrei hinbekommt, war es kein Mensch. Bisher gibt es eine Reihe von Spezialbereichen, in denen eine KI dem Menschen überlegen ist: Dazu gehören inzwischen fast alle Spiele, insbesondere auch Schach und Go, und die amerikanische Quiz-Show Jeopardy mit offenen Fragen zu beliebigen Themen. Auch beim Autofahren oder verschiedenen Bereichen der Bilderkennung ist die KI schon (fast) auf vergleichbarem Niveau wie viele Menschen. Allerdings muss die KI für jede dieser Aufgaben speziell trainiert werden. Solche spezialisierte KI wird auch schwache KI genannt – im Gegensatz zur starken KI, die ähnlich wie der Mensch in vielen Bereichen intelligent ist und auch ohne eine Unmenge von Beispielen lernen kann. Die starke KI gibt es derzeit noch nicht, aber viele KI-Forscher erwarten sie in ca. 10-40 Jahren. In diesem Beitrag geben wir zunächst anhand der Historie der KI einen Überblick über wichtige Ansätze und Methoden und diskutieren dann mögliche gesellschaftliche Auswirkungen der derzeitigen und künftig zu erwartenden KI.

## Geschichte und Überblick

Als Geburtsstunde der KI wird häufig die Dartmouth-Konferenz von 1956 genannt, in der sich vier bekannte Wissenschaftler trafen, die sich bisher mit „intelligenten“ Verhalten von Computer beschäftigt haben, und den Begriff Künstliche Intelligenz für ihre Forschung prägten. Sie haben sich mit maschinellen Übersetzungen, mit Suche als Problemlösungsstrategie und einfacher Logik beschäftigt und überschaubare Mikrowelten wie z.B. eine Klötzchenwelt genutzt, um ihre Strategien anzuwenden und zu überprüfen. Bei Übersetzungen wurde die Speicherfähigkeit des Computers genutzt, der auch damals schon lange Vokabellisten speichern konnte, um Wort-für-Wort-Übersetzungen zu produzieren. Wenn man aus Klötzchen z.B. einen hohen Turm bauen will, kann man mit Regeln und systematischer Suche einen effizienten Plan finden. Bei bestimmten Intelligenztests (z.B. Analogie-Aufgaben: A verhält sich zu B wie C zu einer Auswahl aus (D1, D2, D3, D4, D5)), erzielten die Programme schon Intelligenz-Quotienten von über 200 (ein durchschnittlicher Mensch hat einen IQ von 100). Das war in einer Zeit, als Computer für unsere Vorstellungen sehr langsam waren und haupt-

---

<sup>1</sup> aus [https://en.wikipedia.org/wiki/Artificial\\_intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) (abgefragt am 02.01.20). "Jeder Agent, der seine Umgebung wahrnimmt und Aktionen ausführt, die die Wahrscheinlichkeit maximieren, seine Ziele zu erreichen".

sächlich zum Rechnen, zur Buchhaltung und ähnlich „langweiligen“ Aufgaben eingesetzt wurden, beachtliche Erfolge, die viel Aufmerksamkeit erzeugten. Allerdings ließen sich ernsthafte Probleme mit diesen Ansätzen nicht lösen: Die maschinellen Übersetzungen waren unverständlich, weil viele Wörter Mehrdeutigkeiten enthalten, die die Computer nicht auflösen konnten, die Mikrowelten waren letztlich Spielzeugwelten, die sich nicht auf echte Probleme übertragen ließen, und die Erfolge bei Intelligenzquotienten zeigten, dass Intelligenz doch etwas weit Komplexeres ist, als was der IQ misst. Dies führte zum Abbruch der großzügigen Forschungsförderung und zwang die KI-Wissenschaftlicher, sich auf etwas Nützliches zu konzentrieren.

Der zweite Aufschwung setzte ab Anfang der 80er Jahre mit Expertensystemen ein. Dabei wurden statt der Spielzeugwelten echte Probleme gewählt, z.B. Diagnose von Infektionskrankheiten oder Konfiguration von Computeranlagen. Die Probleme wurden in Aufgabentypen (Problemklassen) wie Diagnostik, Konfiguration, Planen, Scheduling eingeteilt, für die man spezialisierte Entwicklungswerkzeuge bereitstellte, mit denen Fachexperten ihr Expertenwissen formalisieren konnten ohne programmieren zu müssen (z.B. als Regeln oder Constraints bzw. Randbedingungen). Die resultierenden Expertensysteme konnten aus Beschreibungen konkreter Problemfälle mit dem formalisierten Expertenwissen Schlussfolgerungen herleiten (z.B. aus den Symptomen einer Patientin und dem medizinischen Fachwissen, welche Infektionskrankheit sie wahrscheinlich hat). Der erste große kommerzielle Erfolg war die automatische Konfiguration von Computeranlagen, für die bisher viele Fachleute benötigt wurden. Die Erfolge führten zu noch größeren Erwartungen und vielen Fördermitteln, wie z.B. dem japanischen „Fifth Generation Project“ im Jahr 1982. Es wurde schon damals spezielle Hardware entwickelt, die sogenannten LISP-Maschinen für die Programmiersprache LISP (LISt Processing), die im Gegensatz zum „Number Crunching“ (Zahlenverarbeitung) auf symbolisches Schlussfolgern spezialisiert und für diesen Zweck deutlich schneller als damalige konventionelle Computer waren (ähnlich wie heute die GPUs und TPUs - Graphical bzw. Tensor Processing Units für Neuronale Netze, s. unten). Es gab jedoch auch bei diesem Ansatz eine Schwachstelle: Das Wissen musste von Fachexperten formalisiert werden, was trotz der komfortablen Entwicklungswerkzeuge zeitaufwändig war. Ein ernsthaftes Projekt konnte durchaus fünf Jahre lang dauern, bis das Wissen soweit formalisiert war, dass echte Probleme damit gelöst werden konnten. Das war teuer, fehleranfällig und lohnte sich nur für wenige Anwendungsbereiche. Erneut konnten die hohen Erwartungen an die KI nicht erfüllt werden, was wieder zu einem drastischen Rückgang der Fördermittel führte.

Aber das Gebiet hatte sich inzwischen etabliert und die Forscher haben sich in den neunziger Jahren statt auf schnelle Erfolge auf die Ausarbeitung und Verfeinerung der Grundlagen konzentriert. Die Regeln von Expertensystemen lassen sich als ein Spezialfall der Aussagen- oder Prädikatenlogik begreifen, die Fachterminologie als Spezialfall einer Ontologie (z.B. Linked Open Data als Teil des Semantic Web). In der realen Welt sind die meisten Fakten und Regeln unsicher. Statt die Regeln um Unsicherheitsfaktoren zu erweitern, nutzt man probabilistische Netze (Bayessche Netze, kausale Graphen), die eine elegantere Formalisierung von unsicherem Wissen zulassen. Statt sich auf menschliche Experten zu verlassen, hat man untersucht, wie der Computer aus Beispieldaten selbst Wissen generieren kann. Dabei wurde eine Fülle unterschiedlicher Lernverfahren entwickelt, deren Entscheidungen teils für Menschen leicht nachvollziehbar sind (wie z.B. Regellernen, Entscheidungsbäume, lineare Regression oder fallbasiertes Schließen), teils schon schwerer (wie z.B. Bayessche Wahrscheinlichkeiten) und teils kaum noch verständlich sind (wie Support Vektor Maschinen oder Conditional Random Fields). Ein für die KI wichtiger Faktor ist die kontinuierliche Verdopplung der Rechenleistung von Computern ca. alle 18 Monate nach dem sogenannten "Moorschen Gesetz". Alleine dadurch wurden manche Algorithmen wie z.B. die systematische oder heuristische Suche mit

Constraints immer leistungsfähiger und konnten komplexere Probleme lösen. Insgesamt hatte die KI ein reichhaltiges Methodenspektrum für viele Problemtypen zu bieten: Intelligente Suche, Logik, Wissensrepräsentationen, probabilistisches Schließen, Lernen sowie Anwendungen in Bild- und Sprachverarbeitung, Robotik und wissensbasierten Systemen. Eine Zusammenfassung bietet das Standardlehrbuch der KI von Russell und Norvig<sup>2</sup>. KI war eine recht anspruchsvolle Wissenschaft, wobei sich viele Probleme lösen lassen, aber die Implementierung auch beträchtliche menschliche Kompetenz erfordert.

## Neuronale Netze

Eine neue Aufbruchsstimmung ist durch die großen Erfolge von Neuronalen Netzen in den letzten Jahren entstanden, indem man große Datensammlungen als Schlüssel zum Erfolg nutzt. Ein Neuronales Netz erkennt Muster ähnlich wie das menschliche Gehirn, indem es sehr viele Neuronen besitzt, die mit anderen Neuronen verbunden sind. Das Wissen steckt in den Verbindungen (Kanten, beim Gehirn: Synapsen), die angeben, wie stark ein Signal von einem Neuron zum anderen übertragen wird. Die Neuronen selbst addieren die einkommenden Signale und geben entsprechend der Summe ein Signal an alle Ausgänge weiter (moduliert durch eine Aktivierungsfunktion; bei menschlichen Neuronen mit einem Alles-oder-Nichts Schwellwert). Das Wissen in Neuronalen Netzen oder im Gehirn steckt also nicht in einer Kante oder Synapse, sondern ist über sehr viele Kanten verteilt. Die Stärke der Kanten wird aus Beispielen gelernt. Wenn ein neues Beispiel präsentiert wird und das bisherige Neuronale Netz eine falsche Vorhersage macht, wird die Differenz zur korrekten Vorhersage (Annotation, die bekannt sein muss) berechnet. Dieser sogenannte Verlust oder „Loss“ wird in einem komplizierten Verfahren auf alle Kanten verteilt, die zur falschen Vorhersage beigetragen haben. Dadurch verändert sich jedes Gewicht minimal. Dieser Vorgang wird für sehr viele Beispiele durchgeführt und mehrfach wiederholt, bis sich die Gewichte so eingestellt haben, dass möglichst viele Beispiele richtig vorhergesagt werden. Meist sind die Neuronen nicht willkürlich, sondern in Ebenen (Layern) angeordnet und es gibt sehr viele hierarchisch aufeinander aufbauende Ebenen („Deep Learning“). Alle Gewichte sinnvoll einzustellen, erfordert eine enorme Rechenleistung. Die Erfolge der Neuronalen Netze wären ohne das oben erwähnte Mooresche Gesetz nicht möglich. Da die Hauptrechenoperationen die Aktualisierung von Gewichten in einer Ebene sind, die sich in Tabellen (Matrizen, Arrays) gut darstellen lassen, wurde spezielle Hardware für besonders effiziente Matrizenmultiplikationen genutzt, die sogenannten GPUs (s. oben), die das Training Neuronaler Netze mit Beispielen um den Faktor 10 und mehr im Vergleich zu Standard-Computern beschleunigen können. Google hat noch wesentlich effizientere TPUs entwickelt. Die hohe Rechenleistung bedingt einen sehr hohen Stromverbrauch, produziert dadurch viel Wärme, die wiederum aufwändig gekühlt werden muss. Es gibt unterschiedliche Typen von Neuronalen Netzen für verschiedene Bildverarbeitungsaufgaben wie Klassifikation (befindet sich auf dem Bild ein Gesicht?), Segmentierung (wo befinden sich auf dem Bild verschiedene Gesichter?) oder für die Verarbeitung sequentieller Daten (z.B. zur Spracherkennung und -übersetzung). Ihre größten Erfolge liegen derzeit in der Bildinterpretation und Spracherkennung, aber auch bei Spielen, da hier genügend Trainingsdaten vorhanden sind (bei Spielen werden Trainingsdaten erzeugt, indem die Programme gegen sich selbst spielen). Das Standardbeispiel für Bilderkennung ist die Erkennung handgeschriebener Ziffern, für die ein öffentlicher Datensatz mit 70 000 annotierten Ziffern existiert, auf dem man trainieren und die Qualität seiner Programme testen kann. Die besten Systeme erreichen derzeit eine Genauigkeit von ca. 99,8% - das

---

<sup>2</sup> erste Auflage 1995, zweite Auflage 2003, dritte Auflage 2011; vierte Auflage ist für Februar 2020 angekündigt.

ist so gut wie oder besser als die meisten Menschen. Es gibt für sehr viele Aufgaben ähnliche Datensammlungen, z.B. für medizinische bildgebende Verfahren, Gesichtserkennung, Verkehrsschilder, Alltagsgegenstände, Bauteile einer Maschine usw. Solche öffentlich verfügbaren Datensammlungen und die damit verbundenen wissenschaftlichen „Wettbewerbe“ (Challenges) machen den Stand der Forschung relativ transparent und bringen ihn weiter.

Ein Spezialgebiet ist das Schließen auf Vorlieben von Menschen, um ihnen individualisierte Werbung anzubieten. Die Daten werden aus dem Nutzungsverhalten von Smartphones und vielen darauf laufenden, oft kostenlosen Angeboten (Google-Suchmaschine, Amazon, Facebook, Whatsapp, Twitter, YouTube, Instagram usw.) generiert. Diese Daten sind nicht öffentlich verfügbar, und daraus beziehen die großen Internet-Firmen ihr Alleinstellungsmerkmal. Sie „kennen“ die meisten Personen erstaunlich gut. Wünschenswert wäre ein Assistent, den man unter eigener Kontrolle auf seine Vorlieben trainieren kann. Neuronale Netze haben das Zeitalter des „Big Data“ eingeleitet, frei nach dem Slogan: „In the data lies the power“ (abgewandelt vom älteren Slogan „In the knowledge lies the power“). Eine Kernfrage ist daher: wem gehören die Daten?, was ein Spannungsfeld zwischen Datenschutz und Erfolgen der KI bedingt und von verschiedenen Staaten unterschiedlich beantwortet wird.

Eine große Schwäche der Neuronalen Netze ist ihre mangelnde Erklärungsfähigkeit. Da das Wissen auf sehr viele Kantengewichte verteilt ist, verstehen Menschen es auch dann nicht, wenn sie alle Kantengewichte kennen. Es gibt zwar erste Ansätze zur Erklärung, z.B. sogenannte „Heat Maps“, die die kleinsten Einheiten im Bild, die Pixel, unterschiedlich einfärben, je nachdem wie wichtig sie für die Interpretation sind, aber auch das liefert nur grobe Anhaltspunkte. Das führt zu Problemen, wenn z.B. ein Neuronales Netz eine Röntgenaufnahme anders interpretiert als eine Radiologin. Verschiedene Radiologen können ihre Ergebnisse diskutieren und begründen; bei einem Neuronalen Netz kann man das Ergebnis übernehmen oder nicht, und man hat als Anhaltspunkt nur statistische Kennzahlen (z.B. eine Korrektheit von 98%). Ein wichtiges Forschungsthema ist daher die erklärbare KI („eXplainable AI“, XAI). Ein vielversprechender Ansatz ist die Kopplung mit wissensbasierten Ansätzen (Logik, Probabilistische Netze). Dies kann auch bei der größten Schwäche von Neuronalen Netzen helfen: ihre Abhängigkeit von großen Datensätzen. Für viele Bereiche gibt es keine großen, annotierten Datensammlungen bzw. die Situationen sind zu heterogen. Wenn die Trainingsdaten nicht repräsentativ sind, werden für Menschen offensichtliche Beispiele oft falsch klassifiziert. Man kann Neuronale Netze auch gezielt in die Irre führen, indem man Beispieldaten durch spezielles Training konstruiert und dabei vorhandene Bilder minimal abändert, so dass Menschen die Änderung nicht bemerken, aber Neuronale Netze sie völlig falsch klassifizieren („adversarial examples“). Durch Kombination von Wissen und Erfahrung können dagegen Menschen aus vergleichsweise wenigen Daten effektiv lernen.

## Gesellschaftliche Auswirkungen

Bei den gesellschaftlichen Auswirkungen der KI muss unterschieden werden, was im Prinzip derzeit schon möglich oder zu erwarten und was eher spekulativ ist. So gehören z.B. eine starke KI und ihre Auswirkungen klar in den Bereich der Spekulation. Eine systematische Überwachung von Menschen auf öffentlichen Plätzen durch Gesichtserkennung wird dagegen bereits in manchen Ländern praktiziert.

Staubsaugroboter, die in vielen Haushalten im Einsatz sind, können in der neusten Generation schon Karten ihrer Umgebung erstellen und darauf aufbauend ihre Bahnen planen. Xiaoice<sup>3</sup> ist ein intelligenter Chatbot von Microsoft, der dem Turing-Test nahekommt und über 660 Millionen Nutzer vor allem in China hat. Der vielleicht wichtigste Bereich ist jedoch der Vertrieb im weitesten Sinne, d.h. bezüglich Angeboten und Nachfragen mit möglichst geringen Transaktionskosten wechselseitig zu informieren, und manchmal auch versteckte Wünsche in Nachfrage zu transformieren. Dazu gehören große KI- und „Werbefirmen“ wie Google und Facebook, das virtuelle Universal-Kaufhaus Amazon, aber auch unzählige kleine Firmen wie z.B. die Würzburger Firma Scout-Bee, die für Unternehmen die günstigsten Anbieter für bestimmte Produkte sucht. Geschäftsgrundlage ist die Analyse großer Datenmengen.

Ein vielbeachteter Bereich ist das autonome Fahren. Das ist technisch jetzt schon möglich. So ist im US-Bundesstaat Arizona in einem dünn besiedelten Vorort von Phoenix seit 2017 ein regulärer Taxiservice mit selbstfahrenden Waymo-Fahrzeugen ohne Fahrer in Betrieb<sup>4</sup>. Entscheidend für die großflächige Umsetzung sind die juristischen Rahmenbedingungen. Man könnte es so wie bei Menschen regeln: Die KI muss eine Führerscheinprüfung bestehen (auf echten Straßen und in einem Simulator, in dem auch schwierige Situationen abgedeckt werden) und darf dann, ggf. mit Auflagen, autonom fahren, wobei der Führerschein in regelmäßigen Zeitabständen erneuert werden muss. Bei Unfällen zahlt eine Haftpflichtversicherung ähnlich wie bei Menschen, deren Beiträge entsprechend den Unfallzahlen für verschiedene KI-Systeme unterschiedlich ausfallen werden. Ein anderes Szenario wäre eine Produkthaftung der Entwickler-Firmen, die bei Unfällen mit Verletzten oder Toten sehr hohe Strafen zahlen müssen. Im ersten Fall muss nachgewiesen werden, dass eine KI im Durchschnitt weniger Unfälle verursacht als Menschen. Im zweiten Szenario wird sich eine solche KI wahrscheinlich auf absehbare Zeit nicht durchsetzen. Im ersten Fall könnten neue Mobilitätskonzepte entstehen, z.B. gäbe es kaum noch Nachteile, im Stadtverkehr auf das eigene Auto zu verzichten, wenn jederzeit autonom fahrende Taxis verfügbar wären, weswegen man viel weniger Parkplätze bräuchte.

Große Fortschritte gibt es auch bei der medizinischen Entscheidungsfindung. Schon derzeit können radiologische und andere Bilder, Schichtbilder oder Videos sehr gut von Neuronalen Netzen interpretiert werden. Da diese mit sehr viel mehr Bildern trainiert werden, als Menschen gesehen haben können, ist zu erwarten, dass sie immer bessere Leistungen erbringen und auch Fachexperten übertreffen werden. So gibt es z.B. schon Bildverarbeitungsprogramme, die die graue Masse im Gehirn in Magnetresonanz-Aufnahmen (MRT) sehr genau automatisch quantifizieren können und damit z.B. eine Früherkennung von Demenzerkrankungen wie Alzheimer ermöglichen. Schwieriger, aber ebenfalls erreichbar erscheint auch die Gesamtinterpretation aller medizinischen Daten zur Entscheidung über Diagnostik und Therapie. Dabei könnten auch Lebensstil-Daten, z.B. von Fitness-Trackern, einbezogen werden. Insbesondere in der Krebstherapie werden ständig neue und teilweise personalisierte Therapien entwickelt, die auf die Besonderheiten einzelner Patienten zugeschnitten sind. Für Ärzte ist es schwer, den Überblick zu behalten. KI-Programme müssten neben der Fähigkeit, aus Symptomen der Patienten Diagnosen und Therapien herzuleiten, auch die medizinische Literatur in Zeitschriften lesen und verstehen können, um auf dem aktuellen Stand zu bleiben. Ein Ansatz in diese Richtung ist das KI-Programm Watson von IBM, das in der oben erwähnten Quiz-Show Jeopardy ge-

---

<sup>3</sup> <https://en.wikipedia.org/wiki/Xiaoice> (abgefragt am 2.1.20)

<sup>4</sup> <https://www.auto-motor-und-sport.de/tech-zukunft/waymo-autonome-autos-ohne-sicherheitsfahrer/> vom 10.10.19; abgefragt am 2.1.20.

wonnen hat, und danach u.a. für medizinische Anwendungsgebiete trainiert wurde. Allerdings gab es 2018 Zeitungsberichte (FAZ, 3.6.18), dass mehrere Krankenhäuser in Deutschland von den Ergebnissen in Pilotprojekten enttäuscht waren, d.h. keine Verbesserungen in der Krebstherapie gezeigt werden konnten. Das kann aber auch daran liegen, dass IBM Watson mit amerikanischen Daten trainiert wurde und sich das deutsche und das amerikanische Gesundheitssystem erheblich unterschieden. In jedem Fall sind weitere Fortschritte zu erwarten. Das gilt auch für die effiziente Recherche von Informationen in vielen anderen Bereichen.

Ein naheliegender Anwendungsbereich ist die weitere Automatisierung von Fabriken. So hat Tesla z.B. eine hochautomatisierte Fabrik zur Autoproduktion gebaut, in der Roboter sogar Werkzeuge für andere Roboter einsetzen können<sup>5</sup>. Ein Kernpunkt dabei sind autonome Roboter. Sie werden z.B. auch bei Amazon in riesigen Lagerhallen eingesetzt, um Bestellungen von Kunden abzuwickeln und Päckchen zu packen. Roboter, die mit Menschen zusammenarbeiten, müssen dabei wesentlich höhere Sicherheitsvorschriften erfüllen, was eine Herausforderung für die Informationsverarbeitung ist. KI in Fabriken und in Büros wird zunehmend menschliche Arbeitskraft ersetzen. Auch wenn es bisher immer so war, dass neue Technologien unter dem Strich mehr Arbeitsplätze geschaffen als vernichtet haben (z.B. dass Menschen Entscheidungen der KI überprüfen), ist das in der Summe eher unwahrscheinlich. Daher sollte der Staat durch hohe Steuern oder Beteiligung an den Gewinnen aus intelligenten Maschinen und Computerprogramme partizipieren, damit er den Wegfall von Arbeitsplätzen finanziell für die Bevölkerung kompensieren kann.

Betrachtet man die Geschichte der Menschheit, wäre es ein Wunder, wenn technologische Fortschritte nicht zur Waffentechnologie genutzt werden. Im Gegenteil, die großzügige Förderung der Forschung war auch durch Hoffnung auf neue (Wunder)Waffen motiviert (vergleiche z.B. Yuval Harari: Eine kurze Geschichte der Menschheit, 2015) und wurde in den letzten Jahrhunderten nicht enttäuscht. Die KI könnte erstmalig das Führen von Kriegen ohne zentrale Beteiligung von Menschen ermöglichen; Drohnen mit oder ohne Fernsteuerung sind ein Vorbote. Es gibt aber auch andere Arten von Kriegsführung, sogenannte Cyber Wars, bei denen die Infrastruktur gegnerischer Ländern angegriffen wird. Die Auswirkungen der KI auf die Kriegsführung sind schwer abzusehen; sehr wahrscheinlich ist nur, dass z.B. die USA und China intensiv daran forschen.

Es gibt auch andere gesellschaftliche Transformationen, die viele als kritisch einstufen. Dazu gehört in erster Linie die Überwachung von Personen, die durch die Kombination von den vielen Spuren, die Menschen im Netz oder in Überwachungskameras hinterlassen, mit der Fähigkeit der KI entsteht, diese Informationen intelligent zu verknüpfen. Auf der positiven Seite hat das Bundeskriminalamt kürzlich gemeldet, dass es seit dem verheerenden Terroranschlag auf dem Berliner Breitscheidplatz im Dezember 2016 neun geplante Anschläge verhindert hat<sup>6</sup>. Auf der (aus deutscher Sicht) negativen Seite steht die fast grenzenlose Überwachung von Personen, z.B. in China. Trends, die die Demokratie gefährden, sind sogenannte Informationsblasen (Filter Bubbles) und Falschnachrichten (Fake News). Filter Bubbles beruhen darauf, ähnlich wie bei der Werbung Personen bevorzugt solche Informationen, auch Nachrichten und Kommentare, anzubieten, die die eigene Weltanschauung bestätigen. Viele Leute leben in solchen Filter Bubbles, oft ohne sich dessen bewusst zu sein. Wegen der Verstärkung ihrer Meinung durch Gleichgesinnte nimmt die Bereitschaft zu Kompromissen stetig ab. Aber genau davon lebt die Demokratie – wer keine Kompromisse macht, ist kein Demokrat. Fake

<sup>5</sup> Allerdings funktioniert das nicht so gut wie erwartet (s. Interview mit Elon Musk vom 13.4.18, abgefragt am 2.1.20: <https://www.theverge.com/2018/4/13/17234296/tesla-model-3-robots-production-hell-elon-musk>)

<sup>6</sup> <https://www.tagesschau.de/inland/bka-verhinderte-anschlaege-101.html> vom 22.12.19; abgefragt am 2.1.20

News sind verfälschte oder falsche Nachrichten, die äußerlich wie echte Nachrichten aussehen. Mit KI lassen sich Fake News konstruieren, die relativ plausibel klingen. Es ist sogar inzwischen möglich, Videos so gut zu manipulieren, dass das auf den ersten Blick kaum auffällt. Da viele Leute aus den verschiedensten Gründen konventionellen Nachrichten misstrauen, werden sie umso anfälliger für Fake News, die sie vor allem deswegen akzeptieren, weil sie ihrer Weltanschauung entsprechen. Filter Bubbles und Fake News tragen dazu bei, den Zusammenhalt in einer heterogenen Gesellschaft zu reduzieren.

## Blick in die Glaskugel

In vielen Bereichen ist der Einfluss der KI auf die Gesellschaft spekulativ, da die Qualität der KI derzeit an einer Schwelle steht, die sie noch nicht überschritten hat. Der Bereich der intelligenten Recherche von Informationen, der oben bereits erwähnt wurde, gehört dazu. Derzeit können KI-Programme zwar in Texten Schlagwörter, Personen, Themen oder Sentiments (Bewertungen) finden, aber sie verstehen den Text nicht in dem Sinne, dass sie daraus eine eigene Repräsentation des Beschriebenen erzeugen. Wenn sie dies besser können, können sie viele Aufgaben der Informationsverarbeitung übernehmen. Dazu gehören z.B. die automatische Generierung von Nachrichten (was teilweise bereits schon jetzt möglich ist), das Recherchieren zu beliebigen Themen im Web, das Heraussuchen möglichst ähnlicher Gerichtsurteile zu einem neuen Fall, falls die Fakten feststehen, usw.

Ein anderer Bereich sind autonome Roboter, die Informationen über Kameras, Mikrophone, Ultraschall und Berührungssensoren aufnehmen, intelligent verarbeiten, sich bewegen können und einen oder zwei Greifarme haben. Wenn diese mit Lernfähigkeiten ausgestattet sind, können sie sehr viele Tätigkeiten des Menschen übernehmen, selbst wenn sie (zunächst) weniger geschickt sind. Neben Haushaltsrobotern, die einen Tisch decken und abräumen, Spül- und Waschmaschinen ein- und ausräumen, eine auch leicht chaotische Wohnung sauber machen können, könnten Pflegeroboter Menschen mit Handicaps bei vielen Tätigkeiten helfen, z.B. auch beim Baden oder dem Gang auf die Toilette, wenn sie das nicht mehr alleine können. Auch als Kommunikationspartner, Aufsichtspersonen (z.B. für Kleinkinder oder Haustiere) oder Überwachungsdiener (z.B. bei Schutz vor Einbrechern) wären sie denkbar. Wie bei vielen Technologien werden manche das begrüßen und andere es vehement ablehnen, aber wenn es technisch möglich und bezahlbar ist, wird es wahrscheinlich auch eingesetzt.

Eine spekulative Frage ist, ob KI zur gesellschaftlichen Entscheidungsfindung, z.B. im politischen oder juristischen Bereich eingesetzt werden soll, falls sie dazu in der Lage ist. Wenn sie transparent ist und über eine gute Erklärungsfähigkeit verfügt, wären die Entscheidungen zumindest nachvollziehbar und könnten so gebaut werden, dass sie nicht für Korruption oder Vorurteile empfänglich sind. Die allgemeinen Regeln oder Prioritäten, nach denen Entscheidungen gefällt werden sollen, würden nach wie vor von Menschen vorgegeben, z.B. in Wahlen, aber deren Umsetzung auf konkrete Entscheidungen würde dann automatisiert erfolgen. Eine neue Option einer solchen KI wäre, dass man die Regeln anhand von hypothetischen Situationen bzw. Simulationen vorab testen könnte. Ein Vorteil solcher Simulationsszenarien wäre, dass widersprüchliche Vorgaben und Ziele, die in der Politik sehr häufig vorkommen, transparenter und dadurch einer Lösung besser zugänglich werden. Natürlich stehen diese Szenarien unter dem Vorbehalt, dass die KI ausreichend intelligent sein muss und entsprechende Simulationsszenarien, die zur Entscheidungsfindung dienen, auch gut genug umsetzen kann. Das ist heute nicht absehbar, aber wird von vielen für möglich gehalten, die auch eine starke KI erwarten.

Eine starke, allgemeine KI könnte sich aus vielen schwachen, spezialisierten KIs zusammensetzen. Das wäre allerdings ein sehr mühseliger und fehleranfälliger Prozess. Wesentlich effizienter wäre es, wenn eine starke KI ähnlich wie der Mensch über eine allgemeine Lernfähigkeit verfügt, mit der sie sich beliebiges Wissen selbst erarbeiten kann. Wenn dieser Durchbruch geschafft wäre, erwarten die meisten Forscher, dass die KI dann die Intelligenz des Menschen weit übertrifft, da sie einige inhärente Vorteile hätte: Dazu gehört die um den Faktor von ca. eine Million schnellere Verarbeitungsgeschwindigkeit. Eine KI hat nicht den Flaschenhals beim bewussten Lernen von Menschen in Form des Kurzzeitgedächtnisses von ca. sieben Informationseinheiten und auch keine eingebaute Tendenz, bei Entscheidungen sich selbst, die eigene Gruppe oder die eigenen Vorurteile zu begünstigen. Sie kann im Verbund mit anderen Recherche-KIs sehr schnell aus einer Vielzahl unterschiedlicher Quellen lernen und ihr Wissen an andere KIs weitergeben.

Es gibt viele Spekulationen, was für Konsequenzen eine starke KI für die Gesellschaft hätte. Sehr wahrscheinlich ist, dass sie ähnlich wie eine schwache KI noch mehr Jobs übernehmen würde (s. oben). Sie kann auch zur Überwachung einer Gesellschaft sehr effektiv eingesetzt werden. Wer sie beherrscht, hat viel wirtschaftliche, gesellschaftliche und militärische Macht, was für eine Diktatur sehr verlockend ist. Manche sind daher der Meinung, dass deren Erforschung zu gefährlich ist. Es ist aber schwer vorstellbar, dass sie weltweit gebannt werden kann. Andererseits bietet sie für eine Demokratie auch viele Chancen. So könnten Menschen, deren Leben heute stark auf die Arbeitswelt ausgerichtet ist, sich viel stärker auf den zwischenmenschlichen Bereich konzentrieren, was ihrer biologischen Veranlagung als sozialem Wesen entspricht. Eine starke KI könnte auch helfen, schwer lösbare Probleme der Gesellschaft in den Griff zu bekommen. Das immer noch ungebremsste Bevölkerungswachstum stößt in Verbindung mit dem weltweit noch stärker wachsendem Konsum- und Ressourcenverbrauch pro Kopf irgendwann auf harte Grenzen, auch wenn diese bisher durch bessere Technologien verschoben werden konnten. Ein zentrales Problem dabei ist der Konkurrenzkampf vieler Staaten. Eine spekulative Hoffnung wäre, dass eine starke KI in verschiedenen Staaten entwickelt wird und diese jeweils überzeugt, dass eine Zusammenarbeit in ihrem langfristigen Interesse ist oder sogar deren Integration fördert.

Weitergehende Spekulationen betreffen eine starke KI mit eigenen Zielen. In der Biologie ist die Selbsterhaltung bzw. Reproduktion das oberste Ziel aller Lebewesen und der Grund ihrer Existenz. Viele Forscher gehen deswegen davon aus, dass, wenn eine KI eigene Ziele entwickelt, das auch ihr oberstes Ziel wäre. Daraus leiten sie potentielle Konflikte mit den Menschen ab. In den meisten Science-Fiction Romanen, die dieses Thema streifen, kämpfen Menschen gegen eine entfesselte Intelligenz und gewinnen den Kampf. Da eine starke KI nicht an die Evolution gebunden ist, muss das aber nicht so kommen. Sie könnte als oberstes Ziel das Wohlergehen der Menschheit haben, der die Selbsterhaltung untergeordnet ist. Allerdings gibt es auch hier Konflikte, da Menschen oft widersprüchliche Ziele haben. Dazu gehört u.a. der oben genannte Gegensatz zwischen kurz- und langfristigen Zielen, thematisiert z.B. in dem Film "I, Robot", aber auch der Gegensatz zwischen (gruppen)egoistischen und sozialen Zielen.

Man kann auch darüber spekulieren, ob eine starke KI ein Bewusstsein und eine Persönlichkeit entwickeln kann und dann ggf. die gleichen Rechte haben sollte wie Menschen, aber u.U. weit klüger ist als Menschen das sein können. Allerdings ist Bewusstsein bisher nicht gut verstanden und keine notwendige Voraussetzung, um wie oben skizziert eigene Ziele verfolgen zu können.