Wir wissen heute, dass der Sommer nicht ausgereicht hat. Tatsächlich arbeiten wir noch heute an all diesen Problemen.

Im ersten Jahrzehnt nach der Dartmouth-Konferenz konnte die KI mehrere große Erfolge verbuchen, darunter den Algorithmus von Alan Robinson für allgemeine logische Schlussfolgerungen[2] und das Dameprogramm von Arthur Samuel, das sich selbst beibrachte, seinen Schöpfer zu schlagen.[3] Die erste KI-Blase platzte in den späten 1960er-Jahren, da frühe Bemühungen in den Bereichen Machine Learning und maschinelle Übersetzung die in sie gesetzten Erwartungen nicht erfüllten. Ein von der britischen Regierung 1973 in Auftrag gegebener Bericht kam zu folgendem Schluss: »In keinem Bereich des Forschungsgebiets haben die bisherigen Entdeckungen zu den versprochenen gewaltigen Auswirkungen geführt.«[4] Anders ausgedrückt: Die Maschinen waren einfach nicht klug genug.

Mit meinen elf Jahren kannte ich diesen Bericht zum Glück nicht. Als ich zwei Jahre später den programmierbaren Taschenrechner Sinclair Cambridge bekam, wollte ich ihn intelligent machen. Leider konnten Programme für das Gerät maximal 36 Zeichen umfassen, was für KI auf menschlichem Level nicht ausreicht. Das konnte mich nicht aufhalten. Nachdem ich Zugang zu dem riesigen Supercomputer CDC 6600[5] am Imperial College London bekommen hatte, schrieb ich ein Schachprogramm: einen 60 cm hohen Lochkartenstapel. Das Programm selbst war nicht besonders gut, aber das war nicht wichtig. Ich kannte jetzt meine Bestimmung.

Mitte der 1980er war ich Professor in Berkeley und die KI war wieder im Kommen – dank des kommerziellen Potenzials der sogenannten Expertensysteme. Die zweite KI-Blase platzte, als sich diese Systeme für viele der Aufgaben, die sie übernehmen sollten, als unzureichend erwiesen. Auch hier waren die Maschinen nicht klug genug. Ein KI-Winter brach an. Mein eigener KI-Kurs in Berkeley, der heute mit über 900 Studierenden aus allen Nähten platzt, wurde 1990 von lediglich 25 Personen besucht.

Die KI-Community hatte ihre Lektion gelernt: Klüger war in jedem Fall besser. Aber wie ließ sich das erreichen? Die Mathematik sollte es richten. Wir knüpften an bewährte Disziplinen an: Wahrscheinlichkeitsrechnung, Statistik und Kontrolltheorie. Die Saat für die heutigen Fortschritte wurde in jenem KI-Winter gelegt. Dazu gehörten frühe Arbeiten an großen probabilistischen Schlussfolgerungssystemen und

einem Gebiet, das später unter der Bezeichnung *Deep Learning* bekannt wurde.

Ab etwa 2011 ermöglichten Deep-Learning-Techniken dramatische Durchbrüche in der Spracherkennung, der visuellen Objekterkennung und der maschinellen Übersetzung: drei der wichtigsten bisher ungelösten Probleme der KI. Bei einigen Tests bewältigen Maschinen diese Aufgaben heute ebenso gut oder besser als der Mensch. In den Jahren 2016 und 2017 besiegte AlphaGo von DeepMind den ehemaligen Go-Weltmeister Lee Sedol und den amtierenden Meister Ke Jie. Das hatten Fachleute frühestens für das Jahr 2097 erwartet, wenn überhaupt jemals.[6]

Heute lesen wir in den Medien fast täglich von den Fortschritten der KI. Tausende von Start-ups wurden gegründet, von reichlich Risikokapital finanziert. Millionen Studierende absolvieren Onlinekurse zu KI und Machine Learning. Kein Wunder, locken doch Spitzengehälter von mehreren Millionen Dollar. Die Investitionen von Venture-Fonds, Regierungen und Großunternehmen in diesem Bereich betragen zig Milliarden Dollar pro Jahr. Allein in den letzten fünf Jahren wurde das Feld besser mit Finanzmitteln ausgestattet als in seiner gesamten bisherigen Geschichte. Die Projekte, an denen gerade gearbeitet wird, werden im Laufe des nächsten Jahrzehnts höchstwahrscheinlich große Auswirkungen auf unser Leben haben: selbstfahrende Autos zum Beispiel oder intelligente persönliche Assistenten. Die möglichen wirtschaftlichen und sozialen Vorteile der KI sind gewaltig. Das sorgt natürlich für enormen Schwung in der KI-Forschung.

... und was uns noch erwartet

Bedeutet dieser rasante Fortschritt, dass die Maschinen dabei sind, uns zu überholen? Keineswegs. Bevor es superintelligente Maschinen geben wird, sind noch einige Hürden zu überwinden.

Wissenschaftliche Durchbrüche sind generell kaum vorhersagbar. Wie komplex das Thema ist, zeigt die Geschichte eines anderen Forschungsgebiets, das ebenfalls das Potenzial hat, der Zivilisation ein Ende zu bereiten: die Kernphysik.

Anfang des 20. Jahrhunderts gab es wohl keinen bekannteren Kernphysiker als Ernest Rutherford, den Entdecker des Protons, den Mann, der das erste Atom spaltete (siehe Abbildung 1.2 [a]). Wie seine

Kollegen war sich Rutherford bereits seit langer Zeit darüber im Klaren, dass Atomkerne gewaltige Mengen an Energie speichern. Allerdings war die vorherrschende Meinung, dass man diese Energiequelle nicht anzapfen könne.

Am 11. September 1933 hielt die britische Wissenschaftliche Gesellschaft (*British Association for the Advancement of Science*) in Leicester ihre Jahrestagung ab. Lord Rutherford war Sprecher der abendlichen Sitzung. Wie schon zu früheren Gelegenheiten machte er den Anwesenden die Kernenergie betreffend keine großen Hoffnungen: »Jeder, der in der Umwandlung dieser Atome eine Energiequelle sieht, redet Unsinn.«[7] Am nächsten Morgen war Rutherfords Rede in der Londoner *Times* abgedruckt (siehe Abbildung 1.2 [b]).

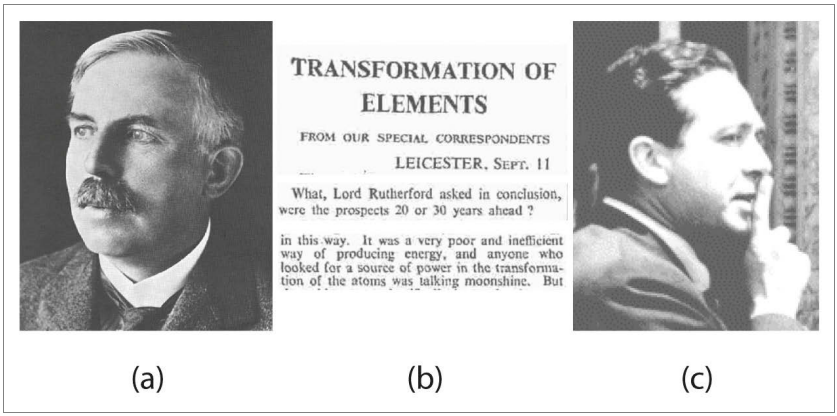

Abb. 1.2: (a) Lord Rutherford, Kernphysiker, (b) Auszug aus einem Bericht der Times vom 12. September 1933 über eine Rede, die Rutherford am Vorabend gehalten hatte, (c) Leó Szilárd, Kernphysiker

Leó Szilárd (siehe Abbildung 1.2 [c]), ein ungarischer Physiker, der kurz zuvor vor dem Nationalsozialismus aus Deutschland geflohen war, wohnte zu der Zeit im Imperial Hotel am Russell Square in London. Beim Frühstück las er den Artikel in der *Times*. Über das Gelesene nachsinnend, unternahm er einen Spaziergang, bei dem ihm die Idee zu der durch Neutronen hervorgerufenen nuklearen Kettenreaktion kam.[8] So wurde das »unmöglich lösbare« Problem der Nutzung der Kernenergie praktisch in weniger als 24 Stunden gelöst. Im folgenden Jahr reichte Szilárd ein geheimes Patent für einen Kernreaktor ein. Das erste Patent für eine Atomwaffe wurde 1939 in Frankreich erteilt.

Und die Moral der Geschichte? Eine Wette gegen den menschlichen Einfallsreichtum ist leichtsinnig, vor allem wenn es um unser aller Zukunft geht. In der KI-Community macht sich eine Art Verleugnungshaltung breit: Manchmal wird bereits die Möglichkeit eines Erfolgs der langfristigen KI-Ziele bestritten. Das erinnert an einen Busfahrer, der die gesamte Menschheit herumkutschiert. In einem Affenzahn hält er auf eine Klippe zu und behauptet gleichzeitig: »Ja klar stürzen wir bei dem Tempo über die Klippe. Aber keine Angst, uns geht vorher noch das Benzin aus!«

Ich sage keinesfalls, dass der Erfolg in der KI *unabdingbar* ist. Und ich bin mir ziemlich sicher, dass es in den nächsten Jahren nicht dazu kommen wird. Dennoch ist es sicher klug, sich auf die Möglichkeit vorzubereiten. Wenn alles gut geht, brechen goldene Zeiten für die Menschheit an, aber wir müssen uns vor Augen halten, dass wir an Dingen arbeiten, die sehr viel mächtiger sind als der Mensch. Wie können wir denn sicherstellen, dass diese Dinge niemals in irgendeiner Weise Macht über uns haben werden?

Damit Sie besser verstehen, wie gefährlich das Ganze ist, hilft ein Blick auf die Auswahl- und Empfehlungsalgorithmen in den sozialen Medien. Sie sind nicht besonders intelligent, und doch beeinflussen sie Milliarden von Menschen und damit indirekt die ganze Welt. Üblicherweise zielen solche Algorithmen darauf ab, die *Klickrate* zu maximieren. Die Klickrate bezeichnet die Wahrscheinlichkeit, mit der ein Nutzer auf die präsentierten Elemente klickt. Wenn wir also einen maximalen Wert erzielen möchten, müssen wir nur solche Elemente anzeigen, auf die Nutzer gern klicken, richtig? Falsch. Die Lösung besteht darin, die Vorlieben der Nutzer so zu verändern, dass sie transparenter werden. Einem besser durchschaubaren Nutzer können Elemente präsentiert werden, auf die er wahrscheinlich klickt und somit mehr Einnahmen generiert. Menschen mit sehr extremen politischen Ansichten sind in dieser Hinsicht meist besser durchschaubar. (Möglicherweise gibt es auch Artikel, auf die extrem gemäßigte Politikinteressierte mit hoher Wahrscheinlichkeit klicken, aber das Thema dieser Artikel ist schwer vorstellbar.) Wie ein rationales Wesen lernt der Algorithmus, wie er den Zustand seiner Umgebung verändern kann, um die eigene Belohnung zu maximieren.[9] In diesem Fall ist die Umgebung die Meinung des Nutzers. Die Folgen sind ein Wiederaufleben des Faschismus, die Kündigung des Gesellschaftsvertrags, auf dem

Demokratien weltweit gründen, und möglicherweise das Ende der Europäischen Union und der NATO. Erstaunlich, was ein paar Zeilen Programmcode anrichten können. Wenn schon ein relativ einfacher Algorithmus dazu imstande ist, wozu ist dann erst ein *wirklich* intelligenter Algorithmus in der Lage?

Was lief schief?

Die Geschichte der KI wird von einem zentralen Mantra bestimmt: »Je intelligenter, desto besser.« Ich bin überzeugt, dass das ein Fehler ist, und zwar nicht, weil ich irgendwie befürchte, ersetzt zu werden, sondern aufgrund unseres Verständnisses von Intelligenz an sich.

Unsere Vorstellung von Intelligenz ist zentral für unser Selbstverständnis. Nicht umsonst bezeichnen wir uns als *Homo sapiens*, als »weiser Mensch«. Nach über 2.000 Jahren der Selbstreflexion lässt sich unsere Idee von Intelligenz wie folgt zusammenfassen:

> Menschen sind insoweit intelligent, als unsere Handlungen darauf ausgerichtet sind, unsere Ziele zu erreichen.

Alle anderen Merkmale für Intelligenz – wahrnehmend, denkend, lernend, erfindend und so weiter – lassen sich als Bestandteile unseres erfolgreichen Handlungsvermögens betrachten. Seit Anbeginn der KI-Forschung wurde Intelligenz in Maschinen ebenso definiert:

> Maschinen sind insoweit intelligent, als ihre Handlungen darauf ausgerichtet sind, ihre Ziele zu erreichen.

Doch weil Maschinen, anders als wir Menschen, keine eigenen Ziele haben, geben wir ihnen diese Ziele vor. Mit anderen Worten: Wir entwickeln Maschinen, die etwas optimieren sollen, geben ihnen Ziele vor und drücken den Startknopf.

Diesen Ansatz finden wir nicht nur auf dem Gebiet der KI. Er zieht sich wie ein roter Faden durch die technologischen und mathematischen Grundsteine unserer Gesellschaft. In der Kontrolltheorie werden Steuerungen für alles Mögliche entwickelt – vom Jumbojet bis zur Insulinpumpe. Dort besteht die Aufgabenstellung für ein System darin, eine *Kostenfunktion* zu minimieren, die angibt, wie stark die Abweichung von einem gewünschten Verhalten ist. In der Wirtschaft gibt es Mechanismen und Richtlinien, die den *Nutzen* Einzelner, das *Wohlergehen* von Gruppen und den *Profit* von Unternehmen