



ALIBABA

IBM

GOOGLE

MICROSOFT

Die großen NEUN

PLASSEN
VERLAG

AMAZON

FACEBOOK

Wie wir die Tech-Titanen
bändigen und eine künstliche
Intelligenz zum Wohle aller
entwickeln können

TENCENT

AMY WEBB

Autorin von *The Signals are Talking*

BAIDU

APPLE

kennen, mit der Automatisierung von Rechenaufgaben. Pascal beobachtete, wie sein Vater mühevoll mit der Hand Steuern berechnete, und wollte ihm die Arbeit erleichtern. Also begann er, an einem automatischen Rechner zu arbeiten, einem mit Zahnrädern und beweglichen Ziffern.⁹ Die Rechenmaschine funktionierte und regte Leibniz dazu an, seine Gedanken zu präzisieren: Maschinen würden zwar nie eine Seele haben, doch es könnte irgendwann möglich sein, eine Maschine zu bauen, die in der Lage sei, so logisch zu denken wie ein Mensch. 1673 beschrieb Leibniz seine Staffelwalzenmaschine, eine neuartige Rechenmaschine, die ein binäres Entscheidungssystem einsetzte.¹⁰ Diese Maschine glich einem Billardtisch mit Kugeln, Löchern, Stöcken und Kanälen. Die Maschine öffnete die Löcher durch Einsatz einer Reihe von Einsen (offen) und Nullen (geschlossen).

Leibniz' theoretische Staffelwalzenmaschine bereitete den Boden für weitere Theorien, darunter die Vorstellung, wenn sich logisches Denken auf Symbole reduzieren ließe und infolgedessen als Rechensystem analysiert werden könnte, und wenn Geometrieaufgaben durch Symbole und Zahlen gelöst werden könnten, sich dann alles auf Bits reduzieren ließe – auch menschliches Verhalten. Damit wich er maßgeblich von den früheren Philosophen ab: Die Maschinen der Zukunft könnten menschliche Denkprozesse wiedergeben, ohne dadurch gegen die göttliche Vorsehung zu verstoßen. Das Denken erforderte nicht unbedingt Wahrnehmung, Sinne oder Seele. Leibniz stellte sich einen Rechner vor, der in der Lage wäre, allgemeine Aufgaben zu lösen – auch solche, die nicht mathematischer Natur wären. Er stellte die Hypothese auf, dass sich Sprache im Rahmen eines universellen Sprachübersetzers auf atomare mathematische und wissenschaftliche Konzepte reduzieren lassen könnte.¹¹

Folgen Geist und Maschine einfach einem Algorithmus?

Wenn Leibniz damit recht hätte, dass Menschen Maschinen mit Seelen wären und eines Tages seelenlose Maschinen erfinden würden, die in der Lage wären, unsagbare komplexe Gedanken zu fassen, dann könnte es auf der Welt eine binäre Klasse von Maschinen geben: wir und sie. Doch damit hatte die Debatte erst begonnen.

Ab 1738 konstruierte der Künstler und Erfinder Jacques de Vaucanson eine Reihe von Automaten für die französische Akademie der Wissenschaften, darunter auch eine komplexe naturgetreue Ente. Sie imitierte nicht nur die Bewegungen einer lebendigen Ente, schlug mit den Flügeln und pickte Körner, sondern konnte auch den Verdauungsprozess nachahmen. Das lieferte den Philosophen neuen Stoff zum Nachdenken: Wenn etwas aussah und quakte wie eine Ente, war es dann wirklich eine Ente? Gehen wir davon aus, dass die Ente irgendeine Art von Seele hat, wäre das dann ein ausreichender Beleg dafür,

dass die Ente auch ein Bewusstsein hat – und alles, was damit zusammenhängt?

Der schottische Philosoph David Hume lehnte die Vorstellung ab, dass die Bestätigung der Existenz als solche bereits der Nachweis eines Bewusstseins sei. Anders als Descartes war Hume Empiriker. Er entwickelte ein neues wissenschaftliches Regelwerk auf der Grundlage beobachtbarer Sachverhalte und logischer Argumente. Während de Vaucanson seine verdauende Ente vorführte – als noch längst keine Rede von künstlicher Intelligenz war –, schrieb Hume in *Ein Traktat über die menschliche Natur*: „Die Vernunft ist nur ein Sklave der Affekte und sollte nichts anderes sein“. Hier verwendete Hume „Affekte“ im Sinne von „nicht rationalen Motivationen“ und wollte sagen, dass unser Verhalten nicht von abstrakter Logik, sondern von solcherlei Reizen gesteuert wird. Sind Eindrücke schlicht unsere Wahrnehmung dessen, was wir sehen, berühren, fühlen, schmecken und riechen können, und Ideen Wahrnehmungen von Dingen, mit denen wir nicht direkt in Kontakt kommen, so glaubte Hume, dass unsere Existenz und unser Verständnis der Welt um uns herum auf einem Konstrukt menschlicher Wahrnehmung beruhte.

Im Zuge der Fortschritte bei der Arbeit an Automaten, die immer realistischer wurden, und ernsthafterer Überlegungen über Computer und denkende Maschinen nahm der französische Arzt und Philosoph Julien Offray de La Mettrie eine ebenso radikale wie skandalöse Untersuchung über Menschen, Tiere und Automaten in Angriff. In einer 1747 zunächst anonym veröffentlichten Abhandlung schrieb La Mettrie, Menschen seien Tieren erstaunlich ähnlich, und ein Affe könne eine menschliche Sprache erlernen, wenn er „richtig ausgebildet“ würde. La Mettrie folgerte ferner, dass Menschen und Tiere nur Maschinen seien, die von ihren Instinkten und Erfahrungen gesteuert würden. „Der menschliche Körper ist eine Maschine, die selbst ihre Triebfedern aufzieht; ... und folgerichtig ist die Seele nur ein Bewegungsprinzip beziehungsweise ein empfindlicher materieller Teil des Gehirns“.¹²

Die Vorstellung, dass Menschen schlicht materiegesteuerte Maschinen seien – Zahnräder, die bestimmte Funktionen ausführten –, implizierte, dass wir weder besonders noch einzigartig waren. Sie deutete ferner darauf hin, dass wir möglicherweise sogar programmierbar waren. Traf das zu, und waren wir bisher bereits in der Lage gewesen, naturgetreue Enten und kleine Mönche zu bauen, sollte daraus folgen, dass sich der Mensch eines Tages selbst nachbauen – und eine Vielzahl intelligenter, denkender Maschinen konstruieren könnte.

War es möglich, eine denkende Maschine zu bauen?

In den 1830er-Jahren experimentierten die Mathematiker, Ingenieure und Wissenschaftler bereits eifrig in der Hoffnung, Maschinen zu bauen, die dieselben Berechnungen

durchführen konnten wie menschliche „Computer“. Die englische Mathematikerin Ada Lovelace und der Wissenschaftler Charles Babbage erfanden die sogenannte „Differenzmaschine“ und postulierten später eine fortschrittlichere „analytische Maschine“, die eine Reihe vorgegebener Schritte einsetzte, um Rechenaufgaben zu lösen. Dass die Maschine womöglich mehr vermochte, als Zahlen zu berechnen, daran hatte Babbage nicht gedacht. Es war Lovelace, die in den Fußnoten zu einer wissenschaftlichen Abhandlung, die sie übersetzte, auf einen brillanten Gedanken verfiel und spekulierte, dass eine leistungsfähigere Version der Maschine auch für andere Zwecke eingesetzt werden könnte.¹³ Konnte die Maschine Symbole manipulieren, die ihrerseits verschiedenen Dingen zugeordnet werden konnten (wie Musiknoten), dann könnte man sie auch für Denkprozesse außerhalb der Mathematik einsetzen. Sie glaubte zwar nicht, dass ein Computer je dazu fähig sein könnte, selbstständig zu denken, doch sie stellte sich ein komplexes System vor, das Anweisungen befolgen und so viele alltägliche Arbeitsabläufe von Menschen nachmachen könnte. Damals erschien das manchem wenig bemerkenswert, doch Ada hatte im Grunde das erste vollständige Computerprogramm für eine künftige, leistungsfähige Maschine geschrieben – und zwar Jahrzehnte vor der Erfindung der Glühbirne.

Knapp 250 Kilometer nördlich von Cambridge, wo Lovelace und Babbage an der Universität arbeiteten, spazierte ein junger mathematischer Autodidakt namens George Boole in Doncaster über ein Feld, als er eine Eingebung hatte und beschloss, sein Leben der Erklärung der Logik des menschlichen Denkens zu widmen.¹⁴ Dieser Spaziergang brachte hervor, was wir heute als Boole'sche Algebra kennen – eine Methode zur Verallgemeinerung logischer Operatoren (wie „und“, „oder“ und „nicht“) durch die Verwendung von Symbolen und Zahlen. So würde sich aus „richtig *und* richtig“ „richtig“ ergeben, was physischen Schaltungen und Gattern im Rechner entspricht. Boole brauchte zwei Jahrzehnte, um seine Ideen formell festzuhalten. Und es sollte weitere 100 Jahre dauern, bis jemand merkte, dass die Boole'sche Logik und Wahrscheinlichkeit dazu beitragen könnte, Computer von automatischen Grundrechnern zu komplexeren Denkmaschinen zu entwickeln. Es war aber noch nicht möglich, eine Denkmaschine zu bauen – die nötigen Prozesse, Werkstoffe und Energie waren noch nicht vorhanden –, sodass die Theorie nicht auf die Probe gestellt werden konnte.

Der Sprung von der theoretischen Denkmaschine zum Rechner, der menschliches Denken nachahmte, vollzog sich in den 1930er-Jahren, als zwei bahnbrechende Arbeiten publiziert wurden: „A Symbolic Analysis of Switching and Relay Circuits“ von Claude Shannon und „On Computable Numbers, with an Application to the *Entscheidungsproblem*“ von Alan Turing. Shannon studierte Elektrotechnik am MIT und belegte als Wahlfach Philosophie – eine ungewöhnliche Wahl. Primäres Referenzmaterial für Shannons Dissertation war Booles *An Investigation of the Laws of Thought*. Sein Doktorvater Vannevar Bush animierte ihn dazu, die Boole'sche Logik auf physikalische Schaltungen zu übertragen. Bush hatte eine weiterentwickelte Version der Analytical Engine – der mechanischen Rechenmaschine von Lovelace und Babbage – gebaut. Der

Prototyp seines Analogrechners hieß „Differential Analyzer“ und war vom Konzept her eher eine Ad-hoc-Lösung. Damals existierte noch keine systematische Theorie, die vorschrieb, wie elektrische Schaltungen auszusehen hatten. Shannons Durchbruch bestand darin, dass er Booles symbolische Logik zur Konstruktion elektrischer Schaltungen heranzog und erklärte, wie diese Logik eingesetzt werden konnte, um mit Einsen und Nullen einen funktionstüchtigen Schaltkreis herzustellen. Shannon hatte herausgefunden, dass Computer zwei Ebenen aufwiesen: die physische (das Gehäuse) und die logische (den Code).

Während Shannon daran arbeitete, physikalische Schaltungen mit Boole'scher Logik zu vereinigen, stellte Turing Leibniz' Universalübersetzer auf die Probe, der sämtliches mathematische und naturwissenschaftliche Wissen darstellen konnte. Turing wollte einen Nachweis zum sogenannten *Entscheidungsproblem* führen. Dieses lautet ungefähr folgendermaßen: Kann es einen Algorithmus geben, der ermitteln kann, ob eine willkürliche mathematische Aussage richtig oder falsch ist? Die Antwort sollte negativ ausfallen. Turing konnte nachweisen, dass kein solcher Algorithmus existiert. Nebenbei erarbeitete er aber ein mathematisches Modell für eine universelle Rechenmaschine.¹⁵

Damit veränderte sich alles. Turing dachte, dass ein Programm und die von diesem herangezogenen Daten in einem Computer gespeichert werden konnten – damals in den 1930er-Jahren ebenfalls eine radikale These. Bis dato ging die Konsensmeinung dahin, dass die Maschine, das Programm und die Daten jeweils unabhängige Einheiten waren. Turings Universalrechner erklärte erstmals, warum alle drei Elemente ineinandergriffen. Aus mechanischer Sicht konnte die Logik, die Stromkreise und Schaltungen zugrunde lag, in dem Programm und den Daten kodiert werden. Diese Aussagen waren von enormer Bedeutung: Das Gehäuse, das Programm und die Daten waren Teil eines Ganzen – vergleichbar mit dem Menschen! Wir sind ebenfalls Gehäuse (unsere Körper), Programme (autonome Zellfunktionen) und Daten (unsere DNA, kombiniert mit indirekten und direkten sensorischen Informationen).

Gleichzeitig kreuzten sich endlich die Wege der langen Tradition der Automaten, die 400 Jahre zuvor mit dem kleinen, gehenden, betenden Mönch begonnen hatte, und der Arbeit Turings und Shannons. Der amerikanische Produzent Westinghouse baute für die Weltausstellung 1939 einen relaisbasierten Roboter namens Elektro the Moto-Man. Es war ein glatter goldfarbener Riese mit Rollen unter den Füßen. Er hatte 48 elektrische Relais, die auf der Grundlage eines Telefon-Relais-Systems funktionierten. Elektro reagierte mittels zuvor aufgezeichneter, über einen Plattenspieler abgespielter Sätze auf Befehle, die über einen Telefonhörer eingegeben wurden. Es handelte sich dabei um einen anthropomorphisierten Computer, der in der Lage war, ohne direkte menschliche Beteiligung in Echtzeit rudimentäre Entscheidungen zu treffen – etwa, was er antworten sollte.

Den Schlagzeilen, Science-Fiction-Kurzgeschichten und Wochenschaunachrichten aus dieser Zeit zufolge hatten die Menschen damit offenbar überhaupt nicht gerechnet und

reagierten schockiert und besorgt auf all diese Entwicklungen. Ihnen kam es so vor, als seien die voll entwickelten „Denkmaschinen“ quasi über Nacht auf der Bildfläche erschienen. Der Science-Fiction-Autor Isaac Asimov veröffentlichte mit „Liar!“ in der Ausgabe von *Astounding Science Fiction* vom Mai 1941 eine visionäre Kurzgeschichte. Er reagierte damit auf die Forschungsergebnisse, die er in Grenzbereichen wahrnahm. Darin trat er für seine drei Gesetze der Robotik ein:

- 1. Ein Roboter darf einem Menschen weder Schaden zufügen noch durch Untätigkeit zulassen, dass ein Mensch zu Schaden kommt.**
- 2. Ein Roboter muss den Befehlen der Menschen gehorchen, außer solchen Befehlen, die ihn in Konflikt mit dem ersten Gesetz bringen.**
- 3. Ein Roboter muss seine Existenz verteidigen, solange er dabei nicht in Konflikt mit dem ersten und zweiten Gesetz gerät.**

Später setzte Asimov noch sein sogenanntes nulltes Gesetz hinzu, das über allen anderen Gesetzen steht: „Ein Roboter darf der Menschheit weder Schaden zufügen noch durch Untätigkeit zulassen, dass die Menschheit zu Schaden kommt.“

Würde eine Denkmaschine wirklich *denken*?

1943 veröffentlichten die Psychiatrieforscher der Universität von Chicago Warren McCulloch und Walter Pitts ihre wegweisende Arbeit „A Logical Calculus of the Ideas Immanent in Nervous Activity“, die ein neuartiges System zur Modellierung biologischer Neuronen in einer einfachen neuronalen Netzwerkarchitektur für Intelligenz beschrieb. Wenn Gehäuse, Programme und Daten, wie Turing behauptet hatte, ineinandergriffen, und wenn der Mensch ein ähnlich elegant konzipiertes Gehäuse mit der Fähigkeit zur Verarbeitung von Daten war, so folgte daraus, dass es möglich sein könnte, eine denkende Maschine zu bauen, wenn diese den für das Denken zuständigen Teilen des Menschen nachempfunden würde – unseren Gehirnen. McCulloch und Pitts postulierten eine moderne Computertheorie des Geistes und des Gehirns, ein „neuronales Netz“. Statt sich auf die Maschine als Hardware und das Programm als Software zu fokussieren, stellten sie sich ein neuartiges symbiotisches System vor, das in der Lage war, enorme Datenmengen aufzunehmen – genau wie wir Menschen. Computer waren damals noch leistungsfähig genug, um diese Theorie auf die Probe zu stellen – doch das Forschungspapier regte andere dazu an, auf ein neuartiges intelligentes Computersystem hinzuarbeiten.

Klarer wurde das Bindeglied zwischen intelligenten Computersystemen und autonomen