

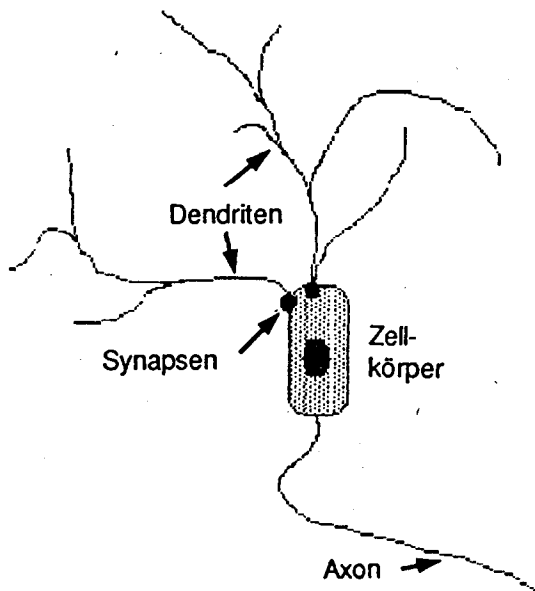
Denkende Maschinen und ihre künstliche Intelligenz

Von Armin P. Barth

Der Traum von der denkenden Maschine zieht sich wie ein roter Faden durch die neuere Geschichte. Einige theoretische Betrachtungen, ein historischer Exkurs und eine aktuelle Bestandaufnahme sollen zeigen, was wir von maschineller Intelligenz oder «Künstlicher Intelligenz» (KI) erwarten dürfen.

Ein Spiel

Wir möchten Ihnen ein Buchstabenspiel erklären. Das Spiel beginnt mit dem Wort «AB». Es sei gleich bemerkt, dass wir nur mit den Buchstaben *A* und *B* spielen dürfen, weshalb die meisten auftretenden Wörter wahrscheinlich bedeu-



Vereinfachte Darstellung eines Neurons.

tungslos sein werden. Wie in jedem Spiel sind nur gewisse Züge erlaubt und andere nicht, was wir in folgenden Spielregeln festlegen:

Regel 1: Wenn Ihnen im Lauf des Spiels ein Wort begegnet, das auf «B» endet, dann dürfen Sie es mit den Buchstaben *B* (von links) und *A* (von rechts) flankieren. (Aus «ABB» würde also beispielsweise «BABBA»).

Regel 2: Ein Wort, das auf «A» endet, darf mit den Buchstaben *A* (von links) und *B* (von rechts) flankiert werden.

Regel 3: «AB» darf man immer zu «BA» umstellen.

Das Ziel des Spieles ist, ein beliebig vorgelegtes Wort, das nur die Buchstaben *A* und *B* enthält, aus dem Startwort «AB» nur durch Anwenden der Regeln 1 bis 3 abzuleiten. Vorgelegt seien nun die Wörter «BABABA» und «ABAAB»; können Sie sie aus «AB» ableiten?

Fall Sie es versucht haben, wird das erste Wort Ihnen keine allzu grossen Schwierigkeiten bereitet haben: aus «AB» leitet man mit Regel 3 «BA» ab, daraus mit Regel 2 «ABAB» und daraus mit Regel 1 «BABABA». Das zweite vorgelegte Wort «ABAAB» hingegen widersetzt sich jedem Lösungsversuch. Vielleicht sind Sie zur Vermutung gelangt, dass es mit Hilfe der Regeln 1 bis 3 aus «AB» gar nicht abgeleitet werden kann. Vielleicht haben Sie sogar entdeckt, woran das liegt: daran nämlich, dass im vorgelegten Wort dreimal der Buchstabe *A* und nur zweimal der Buchstabe *B* erscheint, während Startwort und Regeln aber so beschaffen sind, dass beide Buchstaben stets in gleicher Häufigkeit auftreten müssen.

Formale Systeme

Anhand des obigen Spiels sollte gezeigt werden, was der Mathematiker unter einem formalen System versteht. Ein formales System bearbeitet gewisse in ihrem syntaktischen Aufbau wohldefinierte Zeichenketten (im Spiel waren das die Wörter). Am Anfang stehen ein oder mehrere Axiome, Grundtatsachen des Systems (im Spiel war das «AB»), die entweder keines Beweises bedürfen oder aber von so grundlegender Einfachheit sind, dass sie nicht aus noch Einfacherem abgeleitet werden können. Ein Regelsystem schliesslich formalisiert alle Schlussfolgerungen, die in dem betreffenden System erlaubt sein sollen.

Natürlich erwächst ein formales System nicht einfach der Laune eines Mathematikers, es wird vielmehr mit dem Ziel aufgestellt, einen bestimmten Wirklichkeitsausschnitt möglichst adäquat zu simulieren; die veränderlichen Grössen jenes Realitätsausschnittes finden in gewissen Zeichenketten ihre Darstellung, seine Grundtatsachen (Invarianten) werden durch Axiome eingefangen, seine natürlichen Umbauprozesse schliesslich in Regeln abgebildet.

Ein formales System funktioniert nach seinen eingebauten Spielregeln, unabhängig von allfälligen Interpretationen und kann, gerade dank dieser Losgelöstheit von Herkunft und Umfeldwissen, einer Maschine überantwortet werden. Der Weltausschnitt ist mechanisierbar geworden. Eine Maschine könnte, ebenso wie wir es getan haben, aus dem Axiom «AB» das Wort «BABABA» ableiten, wenn man sie anwiese, auf das Axiom alle Regeln anzuwenden, die anwendbar sind, dann auf die so erhaltenen Zwischenergebnisse wieder alle Regeln anzuwenden, deren Prämissen erfüllt sind, usw. Nach drei Schritten hätte die Maschine eine Menge unnötiger Wörter hergeleitet, aber auch das verlangte «BABABA».

Verlangte man von derselben Maschine aber, das zweite Wort «ABAAB» abzuleiten, so würde

sie in alle Ewigkeit neue Wörter bilden, ohne je zu realisieren (im Sinne einer Bewusstwerdung), dass das fragliche Wort gar nicht beweisbar ist. Die Maschine ist also unfähig, aus dem System, das sie bearbeitet, hinauszuspringen und (öffentlich) Tatsachen über ihre eigene Tätigkeit wahrzunehmen. Können wir also, um eine Hauptfrage der Künstlichen Intelligenz anzusprechen, einer Maschine die Fähigkeit zur *Selbstreflexion* einpflanzen?

Grundfrage der Künstlichen Intelligenz

Eine andere Grundfrage der Künstlichen Intelligenz (KI) lautet: Welcher Teil der Wirklichkeit lässt sich durch formale Systeme modellieren? Die Antwort «nichts» bliebe blind gegenüber vielen Errungenschaften der modernen Computerwissenschaft; die Antwort «alles» würde ein *deterministisches Universum* voraussetzen. Letzteres ist sicher nicht der Fall; zudem wissen wir, dass grosse Teile der Wirklichkeit nicht formalisierbar sind: die Welt ist *chaotisch*.

Verfolgen wir nun einige historische Ansätze der Formalisierung und Mechanisierung von Realitätsausschnitten. Der spanische Denker *Ramon Llull* (1235–1316) beabsichtigte, eine allumfassende und nahezu automatische Methode zur Auffindung von Wahrheiten zu schaffen, die er als «*Ars magna*» bezeichnete. Seine Maschine sollte aus konzentrischen, mit den unterschiedlichsten Begriffen und Zeichen beschrifteten Kreisen bestehen, die, gegeneinander gedreht, *neue Wahrheiten* mechanisch erzeugen sollten. Lulls Gedanken waren sehr wirr und seine Maschine völlig nutzlos. Er hat aber Gedanken vorweggenommen, die in späteren Jahrhunderten, mit einer realistischeren Umsetzung, sehr ernst genommen wurden.

Der englische Philosoph und Staatstheoretiker *Thomas Hobbes* (1588–1679) entwickelte eine Theorie rationeller Erkenntnis in strenger Analogie zum Rechnen. Sowohl die Verknüpfung von Bewusstseinsinhalten im menschlichen Denken wie auch die Verknüpfung der sprachlichen Zeichen, die diese Inhalte ausdrücken, laufen für Hobbes als eine Art Addition oder Subtraktion ab. In Hobbes Worten: «... denn auch Grössen, Körper, Bewegungen, Zeiten, Qualitäten, Handlungen, Begriffe, Verhältnisse, Reden und Namen (worin jegliche Art von Philosophie enthalten ist) können addiert und subtrahiert werden».

Der Wunschtraum des Formalisierens

Im 17. Jahrhundert bauten *Schickard*, *Pascal*, *Leibniz* und andere die ersten Rechenmaschinen. Interessant ist, dass ihre Konstruktion nicht das Diktat der Notwendigkeit war, sondern dass sie den Gehirnen von Philosophen und Theologen entsprangen. Sie verfolgten den Traum, die ganze Welt mit formalen Systemen zu erklären und mit berechnender Vernunft zu zähmen; dieser Traum war vielleicht nur eine frühe Facette des *Aufklärungsgedankens*.

Der Universalgelehrte *Leibniz* (1646–1716) spricht in einem wohl aus dem Jahre 1686 stammenden Projekt davon, von welcher Bedeutung es sein würde, die Prinzipien der Metaphysik, der Physik und der Ethik mit derselben Gewissheit aufstellen zu können wie die Elemente der *Mathematik*. Ein formales System, die «*Mathesis universalis*», sollte Schlussfolgerungen in beliebigen Disziplinen leisten, «derart, dass man seinen Irrtum mit den Augen findet und, wenn es Streitigkeiten unter Leuten gibt, man nur zu sagen braucht: rechnen wir!». Leibniz hat seinen Plan nie verwirklichen können.

Zu Beginn unseres Jahrhunderts, nachdem Mathematiker wie *Boole* und *Frege* eine formale Sprache für das logische Denken geschaffen hatten, rief *David Hilbert* (1862–1943) zu einer Formalisierung der gesamten Mathematik auf. Immerhin versuchte man sich nicht mehr an offensichtlich unformalisierbaren Disziplinen wie *Ethik* und *Metaphysik*, aber die sowieso schon formelhafte (und damals in einer ersten Grundlagenkrise steckende) Mathematik sollte doch formalisierbar sein.

Künstliche Intelligenz heute

Mit der Erfindung des *Computers* wurde eine Mechanisierung formalisierter Abläufe möglich. Und noch einmal erwachte der Traum von der Formalisierbarkeit menschlicher Intelligenz, der 1956 an der Dartmouth-Konferenz den irreführenden Begriff «*Artificial intelligence*» prägte, während die Angst davor sich in Science-Fiction-Romanen wie «*The God Machine*» von *Martin Caidin* niederschlug, in dem ein Elektronengehirn Anspruch auf die Weltherrschaft erhebt. Natürlich wäre es falsch, ein Luddit zu sein: dank ihrer Genauigkeit und Unermüdlichkeit sind Computer für gewisse Aufgaben sehr geeignet, ja sogar *unentbehrlich*. Wir nehmen an, dass sich gewisse menschliche Verhaltensweisen anhand bestimmter Regeln formalisieren lassen und dass diese Regeln von einem Computer verwendet werden können, um das Verhalten zu simulieren.

Betrachten wir nun einige Teildisziplinen der KI; es sind alles nur weitere Versuche, Ausschnitte der Welt formalisierbar und mechanisier-

bar zu machen. Das *Automatic theorem proving* ist ein Spezialfall des Problemlösens. Die Maschine versucht, eine vorgelegte, als Formel in einer bestimmten Syntax beschriebene Behauptung aus den abgespeicherten Axiomen abzuleiten. Sie geht dabei ähnlich vor, wie wir es im AB-Spiel beschrieben haben. Solche Systeme benötigt man nicht nur für das automatische Beweisen mathematischer Sätze, sondern auch beim *automatischen Programmieren* oder beim Entwurf logischer Schaltungen.

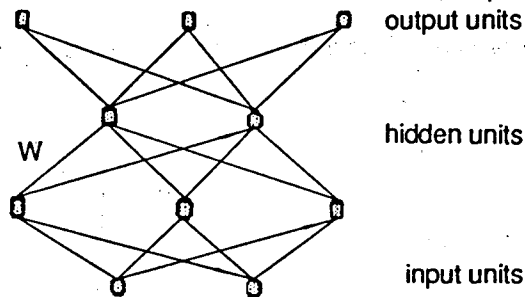
Die Disziplin *Vision and robotics* befasst sich mit Bildanalyse und Bewegung. Sicher wird die Bewegung von Robotern nie so ausgereift sein, dass man etwa Tanzabende für Roboter veranstalten kann. Roboter sind «stark» in der industriellen Fertigung, wo sie Bewegungsabläufe wiederholen, die ihnen vorher in einer *Teach-in-Phase* einprogrammiert wurden. Zudem deuten neuste Forschungsprojekte (wie etwa das *Insect Lab* am MIT) darauf hin, dass ein Paradigmawechsel im Gange ist. Bisher hielt man es für unerlässlich, dass der Roboter zuerst ein Verständnis der ihn umgebenden Szene «berechnet», bevor er sich in Bewegung setzt. Heute baut man Roboter, die weder planen noch ein internes Modell der Aussenwelt aufbauen. Mit einfachsten Versuch-Irrtum-Strategien ausgerüstet, versuchen sie, Raumkenntnisse durch *lernendes Herumkriechen* selber zu erarbeiten.

Expert systems bieten in eng umrissenen Problembereichen Entscheidungshilfen, die den (abgespeicherten) Erfahrungen menschlicher Experten nachgebildet werden. Sie bestehen aus einem *Working memory* (wm), das alle gegenwärtigen Grössen des beteiligten Umfeldes enthält, einer Menge von Regeln der Form *IF «condition in wm» THEN «action»* und einem *Rule-interpreter*, der entscheidet, welche Regel als nächste zur Anwendung gelangt. Solche Systeme erleichtern Diagnose und Therapie in der Medizin, erteilen Expertenrat bei Maschinenreparaturen, Flugzeugreservationsen, Managemententscheidungen usw.

Selbstlernende Netze

Auf lange Sicht am erfolgreichsten wird wohl die Disziplin der *Neural networks* sein. Sie wurden schon vor über 40 Jahren untersucht, dann aber fiel die Forschung in einen Tiefschlaf, weil *Minsky* (1969) in einem Buch (scheinbare) Unzulänglichkeiten aufzeigte. 1981 allerdings bezeichnete er selber sein Buch als «Overkill», und seither erlebt die KI einen beispiellosen Innovationschub auf dem Gebiet der neuronalen Netzwerke; bereits gibt es Firmen, die sich ausschliesslich ihrer Entwicklung widmen.

Neuronale Netze werden der *menschlichen Gehirnstruktur* nachgebildet. Man nimmt an, dass das Gehirn aus zehn Milliarden Nervenzellen (Neuronen) besteht. Über Dendriten werden jedem Neuron die Signale der umgebenden Neuronen in Form eines elektrischen Potentials zugeleitet. Überschreitet die Summe dieser Eingangssignale einen gewissen Schwellenwert (der an den Synapsen «eingestellt» ist), so leitet das Neuron das Signal weiter über das Axon, das zu mehreren tausend Zielneuronen hinführt. Ein neuronales



Schema eines neuronalen Netzwerks. (Bilder Barth)

Netzwerk hat eine ähnliche Struktur. Ein Eingangsmuster wird hochparallel zu einem Ausgangsmuster verarbeitet, wobei die Aktivität jedes Neurons von der Stärke (dem Gewicht) der einzelnen Verbindungen abhängt. Die Gewichte w der einzelnen Verbindungen simulieren die Synapsenstärken und können vom Netzwerk durch geeignete Algorithmen selbständig gelernt werden.

Diese *selbstlernenden Netze* sind wohl nicht zuletzt deshalb so erfolgreich, weil man von den hohen kognitiven Zielen früherer Forschungen abgesehen hat und nur noch versucht, einfaches Handeln in der Umwelt zu simulieren. Bereits sind neuronale Netze im Einsatz bei Mustererkennung, lernenden Robotern, vollautomatischen Qualitätsprüfungen, Echtheitsprüfungen bei Unterschriften und Trendmeldungen im Börsengeschehen.

Grenzen des Formalisierens

Der Formalisierbarkeit sind *prinzipielle Schranken* gesetzt. Zum einen beweisen mathematische Sätze wie die Unvollständigkeitssätze von *Gödel* und *Church*, das Wahrheitstheorem von *Tarski* und andere, dass es immer Wahrheiten geben wird, die ein bestimmtes formales System nicht ableiten kann, so versiert und schnell es auch arbeitet. Die ganze Wahrheit kann also nicht formalisiert werden, und es gibt Erkenntnisse, die nicht auf mechanischem Weg simuliert werden können. Zum anderen besagen die obigen Sätze, dass eine Maschine prinzipiell unfähig ist, gewisse Fragen *über sich selbst* zu beantworten. Dies schliesst die Möglichkeit einer weitgehenden maschinellen Selbstreflexion aus. Wir können nicht formalisieren und *gleichzeitig* den formalisierten Weltausschnitt vollständig erfassen; jedes formale System hat seinen toten Winkel. Der Traum von der denkenden Maschine zieht sich

wie ein roter Faden durch die neuere Geschichte, doch je tiefer wir in formale Systeme und Maschinen eindringen, desto mehr verblasst dieser Traum, desto weniger wird er auch *wünschenswert*. Wir wollen keine Maschinen, die mit uns unsere *Fehler* und *Launen* teilen. Vielmehr wollen wir Maschinen, die uns Arbeiten abnehmen, zu denen wir selbst schlecht geeignet sind. Wohl verlieren wir heute einen Traum, aber wir gewinnen ein *Stauen* über uns selbst.

Adresse des Autors: Balgriststrasse 62, 8008 Zürich.