

Lernen und Adaption	108
Neuronale Netze	114

ANHANG

Glossar	121
Literaturhinweise	128

Einführung

Wer kennt nicht HAL, den sprechenden Computer aus Arthur C. Clarkes und Stanley Kubricks »2001 – Odyssee im Weltraum« aus dem Jahr 1968? Seine Aufgabe ist die Überwachung und Steuerung eines komplexen technischen Systems – des Raumschiffs einer Jupiterexpedition – und dabei insbesondere der vitalen Funktionen der Besatzung. HALs hauptsächliche Interaktionsform mit den Menschen ist das gesprochene Wort, und ein etwas genauerer Blick auf die geführten Dialoge zeigt eine erstaunliche Komplexität in Form, Ausdruck und Inhalt. Niemand wird wohl bestreiten, dass es sich bei der Beschreibung von HALs Sprachbeherrschung um eine technische Vision aus den späten 1960er Jahren handelt, deren Merkmale in ihrer Mehrzahl heute nach wie vor als Forschungsziele im Bereich der Sprachdialogsysteme gelten können.

Die meisten Betrachter des Films würden HAL wohl ohne Zögern »Intelligenz« zusprechen und dabei u. a. an seine spontane Sprachfähigkeit und seine Fähigkeit des plausiblen Folgerns, der flexiblen Planung und der konsequenten Ausführung seiner – wenn auch am Ende böartigen – Pläne denken, ihn aber gleichwohl als ein Artefakt ansehen. Einige sind vielleicht sogar ein wenig gerührt, wenn sie sehen, wie er Modul für Modul stillgelegt wird und dabei in immer frühere Stufen seines Spracherwerbs, der offensichtlich am menschlichen orientiert war, zurückfällt. Es fällt nicht nur beim Menschen schwer, eine klare Trennungslinie zwischen Intellekt und Emotionalität zu ziehen, sondern offensichtlich auch in der als machbar unterstellten maschinellen Form. Dies und die Tatsache, dass HAL keinen eigenen »Körper« hat – es sei denn, man lässt das ganze Raumschiff als solchen gelten –, führt zu zwei wichtigen kritischen Fragen an die Künstliche Intelligenz und die Kognitionswissenschaft, die uns noch beschäftigen werden.

Über ein nahezu perfekt dem menschlichen Körper nachgebildetes äußeres Erscheinungsbild verfügen allerdings die Roboter in Steven Spielbergs Spielfilm »Artificial Intelligence« aus dem Jahr 2001. In den Anfangsszenen werden dem Zuschauer Roboter vorgeführt, die in einer künftigen menschlichen Lebenswelt allerlei nützliche Tätigkeiten verrichten. Auch Kinderroboter, die mit der Fähigkeit zu lieben ausgestattet sind, gehören zum Angebot der Herstellerfirma dieser »Mechas«. Wie der Philosoph John Searle in seiner Filmkritik mit dem Titel »Naturidentische Gefühle« (Die Zeit Nr. 37, 6. September 2001) zu Recht bemerkt, könnten arglose Zuschauer nun erwarten, es würde ihnen mittels der Science-Fiction eine neue Entwicklungsstufe künstlicher Intelligenz vorgeführt – was der Titel ja zu verheißen scheint. Schnell stellt sich jedoch herbe Enttäuschung ein, denn dem äußerst mäßigen Plot der Handlung lässt sich mitnichten entnehmen, wieso man diesen menschenähnlichen Robotern Intelligenz, echte Emotionen oder gar Bewusstsein zusprechen sollte.

Ist »Künstliche Intelligenz« Science-Fiction? Oder stecken hinter der – zugegebenermaßen leicht misszuverstehenden – Bezeichnung nicht doch zahlreiche ernsthafte wissenschaftliche Bemühungen? Nun ist das literarische Genre der Science-Fiction sehr viel älter als die aus dem 20. Jahrhundert stammende Bezeichnung, denn in den Literaturen so gut wie aller Hochkulturen findet man Erzählungen von Androiden, künstlichen Menschen. Und es ist nicht von der Hand zu weisen, dass eine Reihe von belletristisch eingekleideten Zukunftsvisionen in der einen oder anderen Weise einer technischen Realisierung zugeführt wurde.

Das Projekt der »Künstlichen Intelligenz« ist der vorläufige Höhepunkt einer langen Entwicklungsgeschichte, die ihren Ursprung in der Antike hat. Hierbei sind vor allem zwei Entwicklungslinien von besonderer Bedeutung: Zum einen zieht sich der Mythos von der Erzeugung künstlicher Menschen, einer zweiten Schöpfung, durch die Jahrhunderte. Zweitens hat die Idee einer Maschinisierung mensch-

licher Verstandesleistungen ihren Ursprung in der kulturellen Entwicklung der Schriftsprachlichkeit, insbesondere des schriftlichen Rechnens, und führte über immer differenziertere Ansätze der Formalisierung schließlich zur Auffassung des »Denkens als Rechenprozedur« (Sybille Krämer), die auch noch für die moderne Kognitionswissenschaft kennzeichnend ist.

Als dritte, moderne Entwicklungslinie kommt die Erfindung des klassischen programmierbaren Universalrechners hinzu, auf die unsere modernen Computer zurückgehen. Sie sind das Medium, mit dem die technische Umsetzung der Idee des Denkens als Rechenprozedur in Angriff genommen werden konnte.

Es ist alles andere als ein Zufall, dass prominente Pioniere der formalen Logik und der mathematischen Grundlagenforschung sich auch mit dem Bau von Rechenmaschinen befasst haben. Exemplarisch seien hier einige wenige genannt. Gottfried Wilhelm Leibniz drückte im 17. Jahrhundert mit seinem »Calculamus« die Erwartung aus, dass künftige wissenschaftliche Kontroversen durch formales Rechnen in Kalkülen entschieden werden könnten. Zugleich entwarf er eine mechanische Rechenmaschine, die technisch allen Vorgängermodellen überlegen war und kümmerte sich darum, dass sie auch gebaut wurde. Er erfand das binäre Zahlensystem und dachte über eine weitere, nur mit Null und Eins rechnende Maschine nach. Der englische Mathematiker Alan Turing, der in den 1930er Jahren bahnbrechende Untersuchungen in der mathematischen Grundlagenforschung durchführte und einen formalen Begriff der Berechenbarkeit anhand eines sehr einfachen universellen Rechenautomaten einführte, war in den Bau der ersten britischen Elektronenrechner involviert. Er kann als der erste Gründervater der Disziplin »Künstliche Intelligenz« gelten, denn er veröffentlichte bereits im Jahr 1950 einen Aufsatz, ja geradezu ein Manifest mit dem Titel »Computing Machinery and Intelligence«. In ihm schlägt er zur Beantwortung der Frage, »ob Maschinen denken können«, ein Dialogspiel vor, das dann unter

dem Namen »Turing-Test« bekannt wurde. Von der Seite der Rechen-technik herkommend sind u. a. im 19. Jahrhundert der englische Mathematiker Charles Babbage zu nennen und im 20. Jahrhundert der deutsche Computerpionier Konrad Zuse. Zwar konnte Babbages programmierbare Rechenmaschine, die »Analytical Engine«, aufgrund mechanischer Probleme nicht fertiggestellt werden, aber er hatte aufgrund seiner Überlegungen zur arbeitsteiligen industriellen Produktion die Behauptung aufgestellt, dass sich das Prinzip der Arbeitsteiligkeit in gleicher Weise auch auf die geistige Arbeit anwenden ließe. Konrad Zuse, dessen Motivation zur Konstruktion seiner Rechner in der Automatisierung umfangreicher numerischer Berechnungen im Bauingenieurwesen lag, sah im speicherprogrammierbaren Universalrechner ein Werkzeug mit einem weit darüber hinausgehenden Potential; unter anderem beschäftigte er sich mit der Programmierung des Schachspiels, dem in der Frühphase der Künstlichen Intelligenz (KI) durchaus eine Schlüsselrolle zukommt: Sollte einem Computer, der den Schachweltmeister besiegen kann, nicht Intelligenz zugesprochen werden können?

Worin besteht nun das Besondere der KI? Gibt es ein zentrales Thema? Was zeichnet ihre Fragestellungen und Forschungsansätze aus? Und hat die KI eine besondere Methodik?

Die Suche nach einer Antwort auf diese Fragen ist nicht einfach. Russell und Norvig²¹ jedoch bieten eine gute Übersicht. Zusammenfassend kann festgehalten werden, dass sich KI mit Systemen befasst, die

- wie Menschen denken,
- wie Menschen handeln,
- rational denken und
- rational handeln.

Mit anderen Worten: KI ist eine wissenschaftliche Disziplin, die ihre Aufgabe darin sieht, zum einen kognitive Systeme zu simulieren (rationales / menschenähnliches Denken), und zum anderen »intelli-

gente« Systeme zu konstruieren (rationales / menschenähnliches Handeln). Beiden Bestrebungen ist die Erforschung der Prinzipien rationaler bzw. menschenähnlicher Informationsverarbeitung gemeinsam, wobei hier »Information« zunächst in einem alltagssprachlichen, inhaltlichen Sinn gemeint ist. Für diesen Zweck dient als zentrales technisches Medium der Computer, der eine strikte Formalisierung voraussetzt, aber auch exemplarische Realisierungen durch ihre Implementation ermöglicht.

Damit können wir als wohl weitgehend unbestrittenen Kern aller Bemühungen um KI das Ziel verstehen, Artefakte – Computerprogramme oder Roboter – zu bauen, die von Menschen vorgegebene Ziele in einer rationalen und menschenähnlichen und damit nachvollziehbaren Weise erfüllen können. Dieses funktionale, ja nachgeordnete instrumentelle Verständnis des Computers als Werkzeug liegt aber auch der Informatik im Ganzen zugrunde. In ihrem Zugang und ihren Methoden geht die KI aber über die »normale« Informatik hinaus, indem sie in Bereiche vorstößt, die nicht von den Standardverfahren der Informatik erfasst werden. Als Besonderheit mag man vermerken, dass für ihre Herangehensweise die Simulation menschlicher Wahrnehmungs- und Verstandesleistungen eine entscheidende Rolle spielt. Zugleich wird damit auch der grundsätzlich interdisziplinäre Charakter der KI deutlich, der sie zu einer Querschnittsdisziplin zwischen Technik, Naturwissenschaft und Kulturwissenschaften macht: Obwohl durch ihre Entstehungsgeschichte zumeist in der Informatik als Teilgebiet verankert und damit ihre ingenieurwissenschaftliche Komponente betonend, ist KI-Forschung gleichwohl nur in enger Zusammenarbeit mit Philosophie, Psychologie, Linguistik und den Neurowissenschaften möglich, die für ihre kognitionswissenschaftliche Komponente grundlegend sind. Die KI wendet selbstverständlich Methoden der Informatik an und ergänzt sie um heuristische Verfahren, sie befasst sich aber im Unterschied zu den übrigen Zweigen der Informatik mit Problemen, die (noch) nicht in

der dort üblichen Präzision formuliert sind und für die es keine allgemein anerkannten Lösungen gibt. Folglich sind ihre Grenzen fließend und ihre Problemfelder entwickeln sich ständig weiter. Die Entwicklung, Simulation und Überprüfung heuristischer Verfahren ist der zentrale Arbeitsbereich der KI. Mit der Entwicklung und Erprobung ihrer Heuristiken tastet sie sich über Versuche und fortschreitende Verfeinerungen voran, ohne gewiss sein zu können oder gar den Anspruch zu erheben, damit umfassende Problemlösungen zu erreichen. Ein entscheidender Aspekt, warum der Computer zum zentralen Mittel der Forschung werden konnte und damit KI erst möglich wurde, liegt in der grundlegenden Neuerung, dass er ein Maschinentyp ist, der Sprache verwendet. Auch wenn es sich dabei erst einmal um formale Sprachen zur Programmierung von Automaten handelt, stand schon immer die Absicht dahinter, die Kluft zwischen umgangssprachlich oder fachsprachlich formulierten Aussagen und ihrer formalsprachlichen Umsetzung möglichst eng zu machen. Die Fähigkeit, kompetent zu sprechen, schließt aber auch das Vermögen zu urteilendem Denken ein. Daher warf die maschinelle Verwendung der Sprache von Anfang an die Frage auf, ob Maschinen nicht im Prinzip auch denken könnten – wobei das Denken aber immer schon als die Fähigkeit galt, die den Menschen auszeichnet.

Diese zugegebenermaßen radikale Einschränkung auf den genannten Kern erscheint aber nicht wenigen KI-Forschern ungenügend. Dies trifft umso mehr auf das gängige Selbstverständnis der Kognitionswissenschaft zu, die den Computer als Instrument zur Erkenntnis über menschliche kognitive Prozesse nutzen will. Beide möchten die These des »Denkens als Rechenprozedur« über mentalen Repräsentationen absolut setzen in dem Sinn, dass Denken nicht auch als Rechenprozedur verstanden werden kann, sondern als nichts anderes. Mit anderen Worten: Alles Denken ist eine Rechenprozedur. Darin besteht die Kontroverse zwischen der so genannten »schwachen« und der »starken« KI-Hypothese. Während die schwache KI-

Hypothese mit der Werkzeugperspektive auf Systeme der KI verträglich ist, will die starke entscheidend darüber hinausgehen – wobei selbstverständlich das »Alles« das »Auch« einschließt –, und das führt, wie wir noch sehen werden, zu schwerwiegenden philosophischen Problemen.

Wie wir schon an dieser Erörterung feststellen können, ist für den Diskurs in der und über die KI ein extrem metaphorischer Sprachgebrauch mit anthropomorphen Begriffen kennzeichnend. Man kann sie als Metaphern benutzen; wenn wir also von »Denken« bei Maschinen sprechen, muss klar sein, dass wir nicht dasselbe wie menschliches Denken meinen, sondern nur eine Analogie benutzen, in der wir bestimmte Merkmale des menschlichen Denkens ansprechen, die im maschinellen Kontext eine Rolle spielen. Die Benutzung von Metaphern ist bei jungen Disziplinen etwas durchaus Übliches – solange eben die präzisen Fachbegriffe noch nicht erarbeitet sind. Es sollte aber immer klar sein, an welchen Stellen ein Wort als Metapher benutzt wird. Schon die wörtliche Übersetzung von »artificial intelligence« durch »Künstliche Intelligenz« ist nicht unproblematisch, denn sowohl »artificial« als auch »intelligence« haben im Englischen nicht genau dieselbe Bedeutung wie »künstlich« und »Intelligenz« im Deutschen. Um eine etwas grundsätzlichere Auseinandersetzung mit dem Intelligenzbegriff werden wir nicht herumkommen. »Intelligenz ist die allgemeine Fähigkeit eines Individuums, sein Denken bewusst auf neue Forderungen einzustellen; sie ist allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens.« Diese noch recht unpräzise Bestimmung durch den Psychologen William Stern aus dem Jahre 1912 hat eine Vielzahl von weniger überzeugenden Versuchen nach sich gezogen, eine zusammenhängende Intelligenztheorie zu erstellen. Das Spektrum reicht von der bei Stern genannten Anpassungsfähigkeit an neuartige Bedingungen und zur Lösung neuer Probleme bis zu der durchaus ernst gemeinten Feststellung, sie sei eben genau das, was Intelligenztest

messen. Es herrscht breite Übereinstimmung darin, unter Intelligenz »höhere« mentale Prozesse wie Erkenntnisvermögen, abstraktes Denken, Repräsentation, Urteilsfähigkeit, Problemlösen und Entscheidungsfindung zu subsumieren. Diese Fähigkeiten werden aber nicht isoliert, sondern als zusammenhängendes Bündel gesehen, womit dann auch die Annahme eines allgemeinen Intelligenzfaktors begründet wird.

Zweifelsohne wird Intelligenz in besonderer Weise deutlich bei der Fähigkeit, Probleme situationsbezogen zu lösen. Die Art, die Effizienz und die Geschwindigkeit, mit der sich der Mensch bei der Problemlösung an die Umwelt anpasst (Adaption) oder die Umwelt an sich angleicht (Assimilation), ist ein wichtiges Intelligenzmerkmal. In einem hohen Maß spielt dabei die Dynamik der Interaktion mit der Umwelt eine Rolle. Ob allerdings daraus schlüssig die so genannte Emergenzhypothese – dass sich Intelligenz ohne eine angeborene Disposition allein aus dieser Interaktionsdynamik entwickle – begründet werden kann, wie von Vertretern einer neuen Arbeitsrichtung in der KI, der »Nouvelle AI« behauptet wird, stößt doch auf weit verbreitete Skepsis.

Intelligenz äußert sich aber nicht nur in abstrakten gedanklichen Leistungen wie logischem Denken, Rechnen oder allerlei Gedächtnisleistungen und insbesondere in der Fähigkeit zur Reflexion, sondern wird ebenso offenkundig beim Sprachgebrauch oder beim Erkennen von Gegenständen und Situationsverläufen. Neben der konvergen-ten Fähigkeit, eine Vielzahl von Informationen zu kombinieren, um dadurch Lösungen zu finden, spielt bei der Problemlösung aber auch die Kreativität eine wichtige Rolle, insbesondere auch das Vermögen, außerhalb der aktuellen Informationen liegende Lösungsmöglichkeiten einzubeziehen. Andererseits ist gerade die Fähigkeit zur Begrenzung der Suche nach Lösungen bei hartnäckigen Problemen eine typische Leistung der Intelligenz. Von Hirnforschern wurde in der jüngsten Vergangenheit immer deutlicher auf den engen Zusam-

menhang von Intelligenz und Emotionalität hingewiesen. Offensichtlich haben Gefühle wie Lust und Unlust einen deutlichen Einfluss auf die kognitiven Leistungen des Gehirns – sofern sich überhaupt eine klare Trennung vollziehen lässt. Eine Wechselwirkung von Kognition und Emotion scheint sich jedenfalls auch in unserer Alltagserfahrung zu bestätigen, beginnend mit dem Phänomen der aufmerksamen Zuwendung bis zur Erfolgsbewertung unserer Handlungen. Es ist auch bekannt, dass Menschen sich von Artefakten wie Robotern oder Dialogprogrammen, die Emotionalität mit durchaus vordergründigen Mitteln simulieren, erstaunlich schnell anrühren lassen.

Schon wenn wir Intelligenz beurteilen oder gar messen wollen, bedarf es einer Operationalisierung, wodurch wir einen Übergang vom personalen Handeln zum schematischen, nicht-personalen Operieren vollziehen. Das, was operationalisierbar ist, lässt sich grundsätzlich auch mit formalen Systemen darstellen und damit auf einem Computer berechnen. Vieles aber, was das menschliche Denken kennzeichnet, entzieht sich weitgehend einer Operationalisierung. Gerade dies wird jedoch von Vertretern der »starken KI-Hypothese« angezweifelt, die behaupten, dass Bewusstseinsprozesse nichts anderes als Berechnungsprozesse sind, die also Intelligenz und Kognition auf bloße Datenverarbeitung reduzieren.

Wenn wir hingegen Intelligenz erst im sozialen Handlungszusammenhang zuschreiben, können wir dann allerdings auch davon sprechen, dass es – in einem eingeschränkten Sinn – Intelligenz in der Mensch-Maschine-Interaktion gibt, als »Intelligenz für uns«. Es besteht gar keine Notwendigkeit, einem technischen System, das uns als Medium bei Problemlösungen unterstützt, Intelligenz per se zuzuschreiben – die Intelligenz manifestiert sich in der Interaktion. Im Rahmen des instrumentellen Verständnisses erscheint es also weitgehend unproblematisch, von »Intelligenz«, »Emotion«, »Autonomie« etc. zu sprechen: Wenn wir Computerprogramme oder Robo-

ter so bauen, dass sie von Menschen vorgegebene Ziele verfolgen, können wir sagen, dass in ihnen diejenigen Aspekte der Intelligenz realisiert sind, die sie dazu benötigen. Die Simulation bestimmter Formen der Emotion kann dazu dienlich sein, die Mensch-Maschine-Interaktion effizienter zu gestalten. Und wir können ein solches Artefakt »autonom« nennen, wenn es zur Zielerfüllung selbsttätig Unterziele setzen und diese verfolgen kann, also gerade den funktional notwendigen Grad an Autonomie aufweist.

VORGESCHICHTE UND ENTWICKLUNG DER KI

Technische Voraussetzungen

Die *artes mechanicae*, die mechanischen Künste, hatten bereits in der Antike einen erstaunlich hohen Entwicklungsstand erreicht. Auch wenn über die Frühzeit nur wenige bildliche und schriftliche Zeugnisse vorliegen, ist beispielsweise die Errichtung der ägyptischen Pyramiden ohne den Einsatz gewaltiger Baumaschinen nicht erklärbar. Aus der späteren Antike sind jedoch zahlreiche Schriften überliefert, die eine große Vielfalt mechanischer und hydraulischer Maschinen aufzeigen. Zum einen finden sich viele literarische Quellen, in denen sich Sagenhaftes, Phantastisches, Mystisches und Technisches zu Wunderberichten vereinigt. So wird etwa schon aus dem alten Ägypten ca. 2500 v. Chr. von sprechenden Statuen berichtet. Hinweise auf die antike Androidenkunst finden sich in der griechischen Sagenwelt mit der Erschaffung des Talos oder der Pandora. In Homers *Ilias* (800 v. Chr.) erscheinen vom Schmiede-, Handwerks- und Feuergott Hephaistos geschaffene Roboter.

Neben akkuraten Wasseruhren waren zur Zeitmessung und -rechnung offenbar auch Räderwerke in Gebrauch. Ein solches aus dem

ersten vorchristlichen Jahrhundert ist das in einem Schiffswrack gefundene Räderwerk von Antikythera, bei dem gelang, es als astronomischen Computer zu rekonstruieren – ein Maschinenmodell des Kosmos. Aristoteles (384–322 v. Chr.) hatte in seinem Werk die Idee der Welt als vom göttlichen unbewegten Bewegter abhängige Rotationsmaschine dargestellt, und er kommt auch noch in einem anderen Zusammenhang auf Automaten zu sprechen, von denen er offenbar eine durchaus konkrete Vorstellung hatte. In seiner *Politik* entwirft er eine Zukunftsvision von Befehlen gehorchenden und vorausplanenden Werkzeugen, Automaten also, die dem Baumeister die Gehilfen und dem Herrn die Sklaven entbehrlich machen könnten.

Im großen mediterranen Kulturraum wurde neben Mathematik, Naturwissenschaft, Medizin und Philosophie auch das antike technische Erbe von den Muslimen aufgenommen und teilweise signifikant weiterentwickelt, so etwa im Uhrenbau. Automatenfiguren unterhalten die Gäste an den Höfen, und in der arabischen Dichtung gibt es zahlreiche Mythen und Geschichten von Androiden und wunderbaren Automaten. Islamische Wissenschaft und Technik und damit viele bedeutende Werke antiker Autoren fanden vor allem über Spanien und Sizilien den Weg in das lateinische Abendland. Generell gilt für das Mittelalter, dass die technischen Wunderwerke einen ausgesprochen höfischen Charakter haben; ihre Funktionen dienen den militärischen und klerikalen Eliten.

Der allgemeine hohe Stand der spätantiken Technik wurde wohl erst wieder mit der Renaissance erreicht, wiewohl es bemerkenswerte Spitzenleistungen gab, etwa im Uhrenbau mit Giovanni de Dondis Planetenuhr (14. Jh.). Mit der Entstehung der empirischen Naturwissenschaft und der damit einhergehenden Verschmelzung von Wissenschaft und mechanischen Künsten verschiebt sich auch die Perspektive auf die Automaten. Die Untersuchungen des menschlichen Körpers, beginnend mit Leonardo da Vincis anatomischen Zeichnungen, die zum besseren Verständnis der Bewegungsabläufe