Anna Strasser

Kognition künstlicher Systeme

ontos verlag

Frankfurt I Paris I Ebikon I Lancaster I New Brunswick

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie;
detailed bibliographic data is available in the Internet at http://dnb.ddb.de

North and South America by
Transaction Books
Rutgers University
Piscataway, NJ 08854-8042
trans@transactionpub.com

United Kingdom, Ire, Iceland, Turkey, Malta, Portugal by
Gazelle Books Services Limited
White Cross Mills
Hightown
LANCASTER, LA1 4XS
sales@gazellebooks.co.uk

Livraison pour la France et la Belgique:
Librairie Philosophique J.Vrin
6, place de la Sorbonne ; F-75005 PARIS
Tel. +33 (0)1 43 54 03 47 ; Fax +33 (0)1 43 54 48 18
www.vrin.fr

©2006 ontos verlag
P.O. Box 15 41, D-63133 Heusenstamm
www.ontosverlag.com

ISBN 3-937202-96-X

2006

No part of this book may be reproduced, stored in retrieval systems or transmitted
in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise
without written permission from the Publisher, with the exception of any material supplied specifically for the
purpose of being entered and executed on a computer system, for exclusive use of the purchaser of the work

Printed on acid-free paper
ISO-Norm 970-6
FSC-certified (Forest Stewardship Council)
This hardcover binding meets the International Library standard

Printed in Germany
by buch bücher **dd ag**

Danksagung

Bedanken möchte ich mich bei Herrn Prof. Dr. Gerhard Strube für die Betreuung dieser Arbeit und dafür, dass ich diese Arbeit mehrmals im Kolloquium vorstellen und diskutieren konnte. Außerdem danke ich Herrn Prof. Dr. Michael Pauen für die philosophische Betreuung und für seine konstruktive Kritik. Weiterer Dank geht an die DFG, die mich mit einem zweijährigen Stipendium im Rahmen des Graduiertenkollegs *‚Künstliche und Natürliche Intelligen'* gefördert hat.
Und nicht zuletzt möchte ich mich bei den unverzichtbaren KorrekturleserInnen dieser Arbeit, bei meiner Tochter Darja für ihre endlose Geduld mit ihrer gestressten Mama und bei allen Freunden bedanken, die uns in dieser Zeit beigestanden haben. Besonders erwähnen möchte ich Silvia Hermann, die mich in meinen Arbeitspausen immer wieder neu motiviert hat. Ganz besonders dankbar bin ich Regine Amann, die Darja zwei Wochen vor der Abgabe dieser Arbeit mit in den Urlaub genommen hat.
Gewidmet sei diese Arbeit meinen Eltern, Frau Dr. med. Petra und Herrn Dr. med. Florian Strasser, die mir gerade in schwierigen Zeiten immer wieder gezeigt haben, dass man handlungsfähig bleiben kann.

KOGNITION
KÜNSTLICHER SYSTEME

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der Wirtschafts- und Verhaltens-
wissenschaftlichen Fakultät
der Albert-Ludwigs-Universität
Freiburg i. Br.

vorgelegt von

Anna Strasser
aus Freiburg i. Br.

Sommersemester 2004

Inhaltsverzeichnis

Einleitung **7**

I. Was ist Künstliche-Intelligenz-Forschung? **13**
I.1 Geschichte der KI 18
I.2 Paradigmenwechsel in der KI 27
I.3 Die philosophische KI-Debatte 28
I.3.1 Starke KI-These versus schwache KI-These 30
I.3.2 Die Debatte um den Turing Test 33
I.3.3 Die Debatte um das chinesische Zimmer 37
I.3.4 Die phänomenologische Kritik an der KI – Dreyfus 41
I.3.5 Warum Philosophie? 44

II. HANDLUNGEN **47**
II.1 Ereignisklassen 48
II.1.2 Einfaches/komplexes Verhalten versus Quasi-Handlung/ Handlung 56
II.1.3 Handlung als Unterbegriff von Verhalten 70
II.2 Handlungstheorien 72
II.2.1 Georg Henrik von Wright 77
II.2.2 C. G. Hempel 79
II.2.3 Donald Davidson 80
II.2.4 Harry Frankfurt 85
II.2.5 Zusammenfassung 88
II. 3 Kritik am bestehenden Handlungsbegriff 89
II.3.1 Intentionalität 92
II.3.1.1 intrinsische und abgeleitete Intentionalität 97
II.3.1.2 Intentionaler Realismus – eine naturalistische Spielart 99
II.3.1.3 Intentionalität ist Intentionalität ist Intentionalität 99
II.3.2 Speziezismus 100
II.4 Verhältnis Intentionalität – Bewusstsein 101
II.4.1 Zielgerichtetheit 102
II.4.2 Zielgenerierungsfähigkeit 102
II.4.3 Bewusstsein 105

III. Entwicklung des Begriffes einer Quasi-Handlung **107**
III.1 Weltmodell einer dynamischen, komplexen Umwelt 109
III.1.1 Antizipation von Verhalten anderer Systeme 110
III.1.2 Informationsaufnahme für die Bildung eines Weltmodells 112
III.1.4 Handlungskontrolle und -steuerung mithilfe des Weltmodells 113

III.1.5 Planen mithilfe des Weltmodells 114
III.2 Informationsverarbeitung 117
III.2.1 Kognition – Anpassungsfähigkeit – Lernen 120
III.2.1.1 Automatisiertes Verhalten 121
III.2.2 Planen in komplexen, dynamischen Umwelten 127
III.2.3 Kognitive Darstellung der Fähigkeit, Klassisch-Konditionierbar zu sein 128
III.2.4 Menschliche Formen des Lernens in der Psychologie 131
III.2.4.1 Instrumentelles Lernen 138
III.2.4.2 Unterschied: klassisches und operantes Konditionieren 139
III.2.4.3 Formen des Lernens in der KI 140
III.2.5 Erste Beschreibung einer Quasi-Handlung 145
III.2.6 Zwischenresümee 146
II.3 Exkurs: Kognitive Psychologie – Voraussetzungen für Lernfähigkeit 148
III.4 Kognitive Fähigkeiten 152
III.5 Überprüfbarkeit 156
III.6 Wahrnehmung – Informationsaufnahme 159
III.7 Systemeigenschaften der Effektoren bei Quasi-Handlungen 165

IV. Anwendung in der KI 167
IV.1 Definition von Agenten in der KI 168
IV.2 Definition eines autonomen, nicht-kognitiven Agenten 171
IV.3 Definition eines kognitiven Agenten 172
IV.4 Definition eines quasi-handlungsfähigen Agenten 172

V. Quasi-Handlungen 175

VI. Ausblick 177
VI.1 Ethischer Ausblick 177
VI.2 Juristische Konsequenzen 179

VII. Literaturverzeichnis 181
VII.1 Nachschlagewerke im Internet 191

Einleitung

Diese Arbeit ist sowohl aus einem Interesse am Forschungsgebiet der Künstlichen Intelligenz als auch aus der Beschäftigung mit kognitionswissenschaftlicher Theoriebildung entstanden. Im Vordergrund steht das Bemühen, Differenzierungen zwischen verschiedenen Ereignisklassen herauszuarbeiten. Die Ausgangsfrage dieser Arbeit ist, ob man einen Handlungsbegriff vertreten kann, der auch künstliche Systeme als Handelnde zulässt. Handlungsfähigkeit wird in dieser Arbeit als etwas verstanden, das kognitive Fähigkeiten voraussetzt. Ausgehend von einem philosophischen Hintergrund werde ich mich mit der Frage beschäftigen, inwiefern künstlichen Systemen kognitive Fähigkeiten zugeschrieben werden können. Es wird hier nicht die KI-Debatte wieder aufgerollt werden – d. h. es wird nicht diskutiert, ob Maschinen Intentionalität, Selbstbewusstsein oder gar einen freien Willen haben können, sondern es wird eine neue Perspektive auf diese Debatte vorgeschlagen, indem die Frage thematisiert wird, ob künstliche Systeme handlungsfähig sein können, ohne dass man ihnen Intentionalität, Selbstbewusstsein oder einen freien Willen zuschreiben muss. Dabei wird gezeigt, dass es durch eine Erweiterung der Begrifflichkeiten für Ereignisse möglich ist, zwischen verschiedenen Klassen von Ereignissen adäquat zu unterscheiden, die stattfinden, wenn künstliche Systeme involviert sind. Man wird sehen, dass man ohne diese Erweiterung künstlichen Systemen nur Verhalten zuschreiben und nicht weiter differenzieren kann.

Um zwischen verschiedenen Klassen von Ereignissen unterscheiden zu können, werden zum einen die Begrifflichkeiten von ‚Handlung', ‚Kognition' und ‚Lernen' untersucht, zum anderen soll eine Kategorisierung von Informationsverarbeitungsprozessen vorgeschlagen werden.

Der Fokus dieser Arbeit liegt auf den verschiedenen Klassen von Ereignissen, die man beschreiben kann, wobei hier die Klassen ‚Verhalten' und ‚Handlungen' im Vordergrund stehen. In dieser Arbeit wird die Auffassung vertreten, dass der Handlungsbegriff die Möglichkeit einer Ausdifferenzierung der verschiedenen Ereignisklassen bieten sollte. Jedoch ist dies mit dem vorherrschenden traditionellen Handlungsbegriff nicht möglich, denn zumindest aus philosophischer Sicht gilt der Handlungsbegriff ausschließlich für lebendige, meist menschliche Wesen:

> *"The philosophy of action deals with the notion of action that applies only to beings who have wills."* (Ginet 1990, ix)

Handlungsfähigkeit ist zum einen sicherlich eine Fähigkeit, die bei Menschen anzutreffen ist, zum anderen aber ist es sinnvoll zu fragen, inwiefern dieser Begriff auch für Fähigkeiten von komplexen künstlichen Systemen verwendet werden kann, ohne dabei die Unterschiede zwischen Mensch und Maschine verwischen zu wollen. Dafür muss man klären, was die minimalen kognitiven Fähigkeiten für Handlungsfähigkeit sind.

Kognitionswissenschaft setzt voraus, dass es sinnvoll ist, sich auch mit künstlichen Systemen zu beschäftigen. Kann man zeigen, dass künstlichen Systemen Handlungsfähigkeit auf Grund der Art und Weise der Informationsverarbeitungsprozesse zuzuschreiben ist, hat man damit gute Argumente für die Kognitionswissenschaft gesammelt.

In dieser Arbeit werde ich zeigen, wie man den Begriff ‚Handlung' erweitern kann, sodass er nicht a priori auf Ereignisse, die von menschlichen oder wenigstens lebendigen Wesen verursacht sind, einschränkt ist. Hier wird die These vertreten, dass in der Dichotomie der Klassen von Ereignissen, die als Verhalten bezeichnet werden, und derjenigen, die unter den Begriff Handlung fallen, noch differenziert werden sollte. Ausgangspunkt der Argumentation ist das Dilemma, dass es Ereignisse gibt, die offensichtlich komplexer sind als einfaches Verhalten, jedoch den Bedingungen bestehender Handlungstheorien nicht Genüge leisten können. Diese Ereignisse kann man weder der einen noch der anderen Klasse zuordnen. Deswegen wird eine feinkörnigere Unterteilung der verschiedenen Ereignisklassen vorgeschlagen.
So werde ich in dieser Arbeit eine Handlungstheorie entwickeln, mit der man einfache Handlungen (sie werden im Folgenden als Quasi-Handlungen bezeichnet) sowohl von einfachem und komplexem Verhalten als auch von Handlungen im traditionellen Sinn unterscheiden kann. Das Ziel der Arbeit besteht darin, zwischen vier Kategorien von Ereignissen zu unterscheiden:

- **der Klasse von einfachem Verhalten,**
- **der Klasse von komplexem Verhalten,**
- **der Klasse von einfachen Handlungen (Quasi-Handlungen) und**
- **der Klasse von komplexen Handlungen.**

Der Vorteil eines Begriffes einer Quasi-Handlung liegt zum einen darin, dass dieser Begriff nicht von vorneherein auf menschliche Handlungen

eingeschränkt ist, denn er wird keine wesentlich menschlichen Eigenschaften wie z. B. Selbstbewusstsein, Intentionen und freien Willen als Bedingung für Handlungsfähigkeit setzen. Menschliche Handlungen, z. B. ausgeführt mit Selbstbewusstsein, lassen sich als eine Spezialklasse von Verhalten beschreiben.

Zum anderen kann ein Begriff der Quasi-Handlung Unterschiede zwischen verschiedenen Ereignissen, bei denen künstliche Systeme involviert sind, beschreiben. Mit einem Ereignisbegriff, der vom Ereignistyp Handlung menschliche Eigenschaften fordert, kann man Ereignisse, die von künstlichen Systemen verursacht werden, grundsätzlich nur als Verhalten klassifizieren. Jedoch ist es sicherlich unstrittig zu behaupten, dass Informationsverarbeitungsprozesse künstlicher Systeme sich kategorial unterscheiden können. Im Falle einer Quasi-Handlung wird man im Gegensatz zu einem Verhalten eher so genannte kognitive Prozesse erwarten. Diese Arbeit soll zeigen, dass es möglich ist, künstlichen Systemen kognitive Informationsverarbeitungsprozesse zuzuschreiben. Deswegen ist auch eine Klärung des Begriffes ‚Kognition' notwendig. Die Maximalanforderungen, die durch die Zuschreibung von Kognition gefordert werden, sind durch den Menschen erfüllt, jedoch über eine solche Abgrenzung des Begriffes nach unten herrscht keine Einigkeit. Der hier vorgestellte Handlungsbegriff wird eine Abgrenzung des Kognitionsbegriffes vorschlagen. Die Zuschreibung von Kognition geht immer mit der Zuschreibung von Lernfähigkeit einher, sodass auch eine begriffliche Klärung des Lernbegriffes notwendig wird, um den Kognitionsbegriff klarer zu spezifizieren. Es soll dargestellt werden, welche kognitiven Fähigkeiten und welche Form von Lernfähigkeit als Minimalbedingung für die Zuschreibung einer Quasi-Handlung gelten können.

Eine solche Einordnung der Ereignisse, bei denen künstliche Systeme involviert sind, kann weitreichende Konsequenzen mit sich führen. Fragen der Verantwortlichkeit können über die Differenzierung zwischen Verhalten und Handlung entschieden werden. Ein System, welches sich ‚nur' verhält, wird sicherlich als Werkzeug eingestuft, dem man selbst keine Verantwortung für die mit ihm hervorgebrachten Ereignisse zuschreiben kann. Ein System jedoch, dem man Handlungsfähigkeit zuschreiben kann, kann in so einem Fall wohl anders eingeschätzt werden. Die Verantwortung liegt dann wohl eher beim Programmierer (Hersteller oder auch Verkäufer) als beim Benutzer?

Wer ist Schuld, wenn meine Software nicht beim vereinbarten Limit die Aktien verkauft, das ich vorher ausgewählt habe? Wie schwierig ist es in vielen Fällen, eine eindeutige Entscheidung zu treffen, ob man es mit einem Fall von menschlichem oder technischem Versagen zu tun hat?

Fragen wie diese wird meine Arbeit sicherlich nicht beantworten, aber sie kann die Basis für einen Umgang mit solchen Fragen legen. Werkzeuge, die sich nur verhalten, sind nicht verantwortlich, jedoch Werkzeuge, die quasi selber handeln, müssen haftbar gemacht werden. Und da man Programme nicht wirklich bestrafen kann, wird man sich an ihre Erzeuger halten:

„**Eltern haften für ihre Kinder.**“

Diese Arbeit gliedert sich in vier Teile:
In Kapitel I, *Was ist Künstliche-Intelligenz-Forschung?*, wird gezeigt, warum man präzisere Fachbegriffe für Handeln und Verhalten gerade in der Künstlichen-Intelligenz-Forschung (KI-Forschung) braucht. Dies ergibt sich aus der Herleitung einer neuen Perspektive auf die philosophische KI-Debatte, zusätzlich gibt diese Arbeit einen Überblick über die bisherige KI-Debatte. Ein Kritikpunkt an der bisherigen KI-Debatte wird darin bestehen, dass in der philosophischen Debatte nicht genügend zur Kenntnis genommen wurde, worin die Forschungsziele und -erfolge der KI bestehen. Zum anderen werde ich auch die Fokussierung auf die Bereiche des Verstehens, der Intentionalität und des Bewusstseins problematisieren.
Nachdem ich eine Einführung in die wichtigsten Schlüsselbegriffe der Debatte geliefert habe, werde ich in einem zweiten Schritt beschreiben, wie sich die Entwicklung der KI-Forschung selbst vollzogen hat. Dieser knappe Überblick wird illustrieren, warum ich in der KI von einem Paradigmenwechsel spreche. Der Paradigmenwechsel hin zu handelnden Systemen macht es auch aus der Perspektive der KI notwendig, sich erneut mit dem Begriff ‚Handlung’ auseinander zu setzen. Hier zeigt sich, dass es sinnvoll ist, den vollzogenen Paradigmenwechsel in der KI auch aus philosophischer Sicht zur Kenntnis zu nehmen.

Das zweite Kapitel, *Handlungen*, nimmt eine Taxonomie der verschiedenen Ereigniskategorien vor und beschreibt den Forschungsstand bezüglich des Handlungsbegriffes. Um im Hauptteil die Kritik an einem konkreten Beispiel vollziehen zu können, werde ich hier die Handlungstheorie von Donald Davidson ausführlicher referieren. Hierbei steht eine Auseinandersetzung mit Davidson mehr im Hintergrund, stattdessen sollen die Bedingungen, die ich aus Davidsons Handlungstheorie herausarbeiten werde, prototypisch für einen philosophischen Handlungsbegriff stehen.

In Kapitel III, *Entwicklung des Begriffes einer Quasi-Handlung*, werde ich zuerst den bestehenden Handlungsbegriff kritisieren, um dann einen erweiterten Handlungsbegriff vorzuschlagen. Man wird hier zum einen sehen, dass einige Bedingungen nicht notwendig für einen allgemeinen Begriff einer Handlung sind. Sie sind jedoch kennzeichnend für menschliche Handlungen. Zum anderen wird es durch das ‚Weglassen' einiger impliziter Bedingungen notwendig, die übrigen Bedingungen genauer auszuformulieren, da sich sonst der Handlungsbegriff kaum von einem einfachen Verhaltensbegriff unterscheiden lässt. Pointiert kann man sagen, dass diese Arbeit zeigen wird, dass sowohl Thermostate (oder Plattwürmer) nicht als handlungsfähige Systeme eingestuft werden können als auch dass menschliche (lebendige) Wesen nicht die einzige Klasse von möglichen Verursachern von Handlungen bilden.
Im vierten Kapitel, *Anwendungen in der KI*, möchte ich diskutieren, welche Bedingungen eine Agentendefinition fordern muss, damit der beschriebene Agent als möglicher Handelnder in Frage kommen kann. Im Ausblick dieser Arbeit werde ich noch auf juristische und ethische Implikationen der vorgelegten Klassifizierung hinweisen.

I. Was ist Künstliche-Intelligenz-Forschung?

> *„Artificial Intelligence is the science of making machines do things that would require intelligence if done by men."* (M. Minsky, 1968)

Mit dem Begriff ‚Künstliche Intelligenz' (KI) bezeichnet man ein Gebiet der Informatik. Man kann jedoch auch die Informatik als ein Teilgebiet der KI bezeichnen, da zur Forschung im Bereich der KI auch Teilbereiche der Psychologie, Linguistik und Kognitionswissenschaft gehören. Maschinen (Systeme) zu entwickeln, welche Aufgaben bewältigen, für dessen Erledigung ein Mensch Intelligenz benötigt, kann man mit Marvin Minsky als eines der Forschungsziele der KI beschreiben.[1] Jedoch ist das nicht das einzige Ziel der KI, es geht auch darum zu verstehen, was intelligentes Verhalten ist, wie aus folgender Definition aus der ‚*Encyclopedia of Artificial Intelligence*' zu ersehen ist:

> *„Artificial Intelligence is a field of science and engineering concerned with the computational understanding of what is commonly called intelligent behavior, and with the creation of artifacts that exhibit such behavior."* (Shapiro 1992, p. 55)

Eine allgemein gültige Definition von KI ist schwer zu geben. Russell & Norvig (1995)[2] unterscheiden zwischen vier verschiedenen Zielsetzungen in der KI:

---

[1] Vgl.: *„Ability of a machine to perform tasks thought to require human* ***intelligence****. Typical applications include game playing, language translation, expert systems, and robotics. Although pseudo-intelligent machinery dates back to antiquity, the first glimmerings of true* ***intelligence*** *awaited the development of digital computers in the 1940s. AI, or at least the semblance of* ***intelligence****, has developed in parallel with computer processing power, which appears to be the main limiting factor. Early AI projects, such as playing chess and solving mathematical problems, are now seen as trivial compared to visual pattern recognition, complex decision making, and the use of natural language."* (Artificial Intelligence. *Britannica Concise Encyclopedia*. Retrieved August 20, 2004, from Encyclopedia Britannica Premium Service.)

[2] Vgl. Russell & Norvig (1995, p. 5).

Tabelle 1: Zielsetzungen der KI

Systems that think like humans.	*Systems that think rationally.*
Systems that act like humans.	*Systems that act rationally.*

Die erste Zielsetzung – **thinking humanly** – beschreibt das, was man auch als kognitive Modellierung bezeichnen kann, nämlich Denkprozesse so zu modellieren, wie sie beim Menschen ablaufen könnten:

> *„If the program input/output and timing behavior matches human behavior, that is evidence that some of the program's mechanism may also operating in humans."* (Russell & Norvig, 1995, p.6)

Hier wird von den künstlichen Systemen nicht gefordert, dass sie physisch an die Außenwelt angekoppelt sind, das Verhalten der Systeme ist auf sprachlich formulierten Output reduziert. Sie brauchen keinen Körper, da vorausgesetzt wird, dass der Geist mit seinen Denkprozessen unabhängig vom Körper existiert. Die physikalische Simulation einer Person scheint deswegen keine notwendige Bedingung für Intelligenz. Es geht dabei primär um Sprachverhalten (*natural language processing*), Wissensrepräsentation (*knowledge representation*), automatisches Schlussfolgern (*automated reasoning*) und maschinelles Lernen (*machine learning*). Hinzu kommen Erkenntnisse aus der Kognitionswissenschaft und Psychologie.

Acting humanly

Bei dieser Zielsetzung liegt der Fokus auf Verhalten, welches Veränderungen in der Umwelt bewirken kann, deswegen stehen zwei anderen Gebiete der KI mehr im Mittelpunkt: visuelle Wahrnehmung (*vision)* und Robotik (*robotics).*

Sowohl die Wahrnehmung (nicht nur die visuelle) als auch die Umweltveränderungsmöglichkeiten sollen nach dieser Zielsetzung denen von Menschen entsprechen. Oben erwähnte Gebiete spielen natürlich auch hier eine Rolle.

Thinking rationally
Bereiche der KI, die sich dieser Zielsetzung verschrieben haben, haben als Grundlage die Regeln der Logik.[3] Die Informationsverarbeitung folgt hier festgelegten Standards und ist auf Utilitymaximierung ausgerichtet. Rational ist, was effizient zum Ziel führt. Ziel ist nicht – wie bei den am Menschen orientierten Zielsetzungen –, auch menschliche Fehlleistungen zu reproduzieren, sondern diese sollen gerade vermieden werden. Mit Hilfe von Wahrscheinlichkeitstheorie und Entscheidungstheorie kann auch unsicheres, unvollständiges Wissen verarbeitet werden. Jedoch führt das prinzipielle Lösen von Problemen nicht zwangsläufig zur Fähigkeit, Probleme in einer komplexen und dynamischen Umwelt zu lösen. Dieser Ansatz wird auch als *‚The laws of thought approach'* bezeichnet.

Acting rationally
Unter rationalem Handeln versteht man, dass die sich durch eigene Wünsche und Vorstellungen ergebenden Ziele stringent verfolgt werden. Dafür müssen alle Ressourcen effizient eingesetzt werden. Gesucht wird der schnellste Weg zum Ziel, und dies impliziert in den meisten Fällen den Einsatz von Vernunft. Eine andere Bezeichnungsweise für diesen Ansatz ist: *‚The rational agent approach'*.

Schon bei der Beschreibung dieser vier Zielsetzungen der KI-Forschung sieht man, dass im Kontext der KI das Wort ‚Handeln' ohne Einschränkungen benutzt wird. Die eingangs aufgeworfene Frage, ob künstliche Systeme handeln können, scheint innerhalb der KI keine Rolle zu spielen. Selbst wenn es sich um Systeme handelt, die offensichtlich nur einfaches oder komplexes Verhalten zeigen, wird von handelnden Agenten gesprochen.
Aus philosophischer Perspektive wird zwischen **starker und schwacher KI** unterschieden. Pointiert kann man Vertretern der starken KI unterstellen, dass sie Systeme entwickeln wollen, die sowohl so denken als auch so handeln wie Menschen. Solchen Systemen kann man nach ihrer Auffassung Geist – im selben Sinne wie bei Menschen – zuschreiben. Anzumerken ist, dass man, um ein angemessenes Modell des menschlichen Denkens und Handelns zu entwickeln, auch menschliche Fehlleistungen reproduzieren muss. Starke KI reduziert Kognition und Intelligenz auf bloße Datenverarbeitung. Intelligenz muss hier operationalisierbar sein. Vertreter der schwachen KI-These wollen dagegen Systeme entwickeln,

---

[3] Grundlagen der Regeln der Logik gibt es schon bei Aristoteles (vgl. *Organon*, 1831).

die rational handeln und denken. Sie erheben nicht den Anspruch, dass bei diesen Systemen identische Prozesse wie beim Menschen ablaufen. Immerhin halten sie die Entwicklung solcher Systeme für ein gutes Instrument für die Untersuchung des Geistes, betonen dabei aber den Modellcharakter.[4]

Die Schwierigkeit, eine allgemeingültige Definition für die KI zu geben, hat ihren Ursprung in der Ungeklärtheit des Begriffes ‚Intelligenz'. Denn schon die wortwörtliche Übersetzung des Begriffes ‚*Artificial Intelligence*' ist nicht unproblematisch:
Zum einen impliziert ‚*artificial*' im Englischen nicht so sehr, wie das im Deutschen der Fall ist, einen Als-ob-Charakter und zum anderen bedeutet ‚*intelligence*' im Englischen zunächst nur ‚Fähigkeit, Wissen zu erwerben und anzuwenden'[5], während im Deutschen Intelligenz häufig mit der Forderung nach **Bewusstsein** einhergeht:

> *„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."* (Stern, 1912, S. 3, Hervorhebung Verfasserin)

Daneben wird gerade in psychologischen Kontexten die Meinung vertreten, dass Intelligenz das sei, was man mit einem Intelligenztest messen könne.[6] Auch wenn es keine allgemeingültige Definition von Intelligenz gibt, herrscht zumindest eine breite Übereinstimmung darin, unter Intelligenz ‚höhere' **mentale Prozesse** wie Erkenntnisvermögen, abstraktes Denken, Bildung von Repräsentationen, Urteilsfähigkeit, Problemlösen und Entscheidungsfindung zu subsumieren.[7]
Intelligenz äußert sich also in logischem Denken, Rechnen und verschiedenen Gedächtnisleistungen, insbesondere in der Fähigkeit zur Reflexion, zum Sprachgebrauch oder zum Erkennen von Gegenständen und Situationsverläufen. Deswegen fokussiert KI-Forschung auf mögliche

---

[4] Der Unterschied zwischen starker und schwacher KI wird in Kapitel I.3.1 ausführlicher dargestellt.
[5] *„**Intelligence:** the ability to learn or understand or to deal with new or trying situations;* ***artificial intelligence: 1:*** *the capability of a machine to imitate intelligent human behavior* ***2:*** *a branch of computer science dealing with the simulation of intelligent behavior in computers" (*Intelligence. *Merriam-Webster Online Dictionary.* Retrieved August 20, 2004, from http://www.m-w.com/cgi-bin/dictionary).
[6] Vgl. Alfred Binet (1916).
[7] Vgl. Görz & Nebel (2003, S.10).

Komponenten von Intelligenz wie z. B. auf **Lernen**, logisches Schlussfolgern, Problemlösen, **Handeln**, Wahrnehmen und Sprachverstehen. Probleme sollen hier situationsbezogen gelöst werden, d. h. sowohl die Anpassungsfähigkeit an die Umwelt als auch die Fähigkeit, die Umwelt zu verändern, spielen eine wichtige Rolle.
Ob Intelligenz an einen bestimmten Körper (Hardware) gebunden sein muss, ist eine Frage, die letztendlich nicht entschieden werden kann. Die philosophische Grundlage der Anfänge der KI setzt in der Regel eine funktionalistische Sichtweise[8] voraus, bei der eine Unabhängigkeit von der Hardware postuliert wird. In der *Nouvelle AI*, einem neuen Zweig der KI, spielt die Körperlichkeit (*embodiment*) wiederum eine gewichtige Rolle.
Zusammenfassend kann man sagen, dass ein Hauptziel der KI in der Operationalisierung menschlicher Fähigkeiten besteht:

> *„Künstliche Intelligenz ist eine wissenschaftliche Disziplin, die versucht, sowohl menschliche Wahrnehmungsleistungen als auch Verstandesleistungen zu operationalisieren.“* (Görz, 2000, S. 1)

> *„KI ist die Untersuchung von Berechnungsverfahren, die es ermöglichen wahrzunehmen, zu schlussfolgern und zu handeln.“* (Winston, 1987)

Vorausgesetzt wird in obiger Definition die so genannte **Berechnungsthese**. Jedoch haben gerade in letzter Zeit Hirnforscher auf den Zusammenhang von Intelligenz und Emotionalität hingewiesen.[9] Emotionalität wie auch Bewusstsein kennzeichnen menschliches Denken, entziehen sich meiner Auffassung nach bis jetzt hartnäckig der Operationalisierung, womit Grenzen für die KI gesetzt werden.[10]

Betrachtet man die Terminologie der KI, kann man Folgendes feststellen:

> *„Da die KI eine junge Disziplin ist, zeichnet sich ihre Grundlagendiskussion zudem durch eine metaphernreiche und*

---

[8] *„Der Funktionalismus ist eine Lehre über die Natur von Geisteszuständen, derzufolge diese wesentlich dadurch charakterisiert sind, welche kausale (oder funktionale) Rolle sie im Gesamtsystem einer Person (oder eines sonstigen geistbegabten Subjekts) einnehmen.“* (Saporiti, 1996, p. 192).
[9] Vgl. Damasio (2000).
[10] Sollte sich herausstellen, dass man Emotionalität und Bewusstsein doch operationalisieren kann, wäre diese Begrenzung der KI aufgehoben.

> *aufgrund ihres Gegenstandes stark anthropomorphe Sprache aus.“* (Görz, 2000, S. 1)

Die Ausbildung von präzisen Fachbegriffen ist noch nicht erfolgt, wie auch Görz & Nebel feststellen:

> *„Die Benutzung von Metaphern ist bei jungen Disziplinen etwas durchaus Übliches – solange eben die präzisen Fachbegriffe noch nicht erarbeitet sind.“* (Görz & Nebel, 2003, S. 9)

Man darf also die in der KI verwendeten Begriffe nicht wortwörtlich verstehen. Bzw. man muss sie klar definieren.
Nach einem kurzen historischen Überblick werden in diesem Kapitel verschiedene Forschungsgebiete (*Communities*) der KI vorgestellt. Außerdem soll gezeigt werden, dass man in der KI von einem **Paradigmenwechsel** sprechen kann, der es notwendig macht, sich mit dem Handlungsbegriff auseinander zu setzen. Ich bin der Auffassung, dass sich die Ziele in der KI inzwischen verschoben haben, weswegen es sinnvoll ist, die philosophische KI-Debatte aus einer neuen Perspektive heraus wieder zu beleben. Argumente gegen die ‚alte´ KI müssen nicht zwangsläufig für die jetzige KI gelten.

I.1 Geschichte der KI

Als Geburtsstunde der Künstlichen Intelligenz gilt die Konferenz von Dartmouth: 1956 (zwei Jahre nach dem tragischen Selbstmord von A. Turing) organisierten John McCarthy (Mathematiker in Dartmouth), Marvin Minsky (Mathematiker und Neurologe in Harvard), Nathaniel Rochester (Manager bei IBM) und Claude Shannon (Mathematiker bei Bell Telephone Laboratories) in Vermont einen zweimonatigen Workshop mit dem prägenden Titel: ‚*The Dartmouth summer research project on artificial intelligence´*. Im Förderungsantrag wurde Folgendes ausgeführt:

> *„Wir schlagen eine zweimonatige Untersuchung der Künstlichen Intelligenz durch zehn Personen vor, die während des Sommers 1956 am Dartmouth College in Hanover, New Hampshire, durchgeführt werden soll. Die Untersuchung soll auf Grund der Annahme vorgehen, dass jeder Aspekt des Lernens oder jeder anderen Eigenschaft der Intelligenz im Prinzip so genau beschrieben werden kann, dass er mit einer Maschine simuliert werden kann.“* (zitiert nach McCorduck, 1987, S. 93)

Durch diese Konferenz bekam die KI ihren Namen; begonnen hat die KI-Forschung jedoch schon früher: z. B. durch Warren McCulloch & Walter Pitts (1943), die schon 1943 Überlegungen zur Konstruktion künstlicher neuronaler Netze (KNN) veröffentlichten, wie such durch die Arbeiten von Donald Hebb, Donald Michie, Alan Newell, Herbert Simon, Alan Turing, von Neumann und Dean Edmonds.
Auch kann man in der Frühzeit der KI-Entwicklung auf die Kybernetiker Claude Shannon (1948) und Norbert Wiener (1948) verweisen. Einen Ursprung der starken KI–These kann man in Norbert Wieners Beobachtungen über Prinzipien der Feedbacktheorie sehen. Die Beschreibung eines Thermostaten bildet das berühmte Beispiel aus dieser Theorie: N. Wiener behauptet, dass man jedes Verhalten, das aus einem Feedbackmechanismus resultiert, als **intelligentes Verhalten** bezeichnen kann.
Die technischen und auch philosophischen Voraussetzungen für die KI haben eine viel weiter zurückreichende Geschichte.

Philosophische Grundlagen

Philosophische Grundlagen der KI findet man schon bei Sokrates und Platon, die untersuchen, ob Wissen in eindeutigen Definitionen formulierbar ist. Auch der Ursprung der Berechnungsthese, nämlich die These, dass man schlussfolgerndes Denken auf spezielle Rechenverfahren reduzieren kann, ist in dieser Zeit anzusiedeln.
Die Idee, dass Denken prinzipiell operationalisierbar ist, d. h. dass Denken im Grunde ein Berechnungsprozess ist, findet man sowohl bei Boole (1854) als auch bei Hobbes:

> *„Die Römer nannten Geldrechnungen rationes […], und was wir in Rechnungen oder in der Buchführung als Posten bezeichnen, nannten sie nomina, das heißt Namen. Und von da her scheinen sie die Bedeutung des Wortes ratio auf die Fähigkeit des Rechnens in allen anderen Gebieten ausgedehnt zu haben. Die Griechen besitzen für Sprache und Vernunft nur das Wort logos. Nicht, daß sie gedacht hätten, es gebe keine Sprache ohne Vernunft, sondern umgekehrt kein Denken ohne Sprache. Und den Denkakt nannten sie Syllogismus, was Zusammenzählen der Folgerungen aus einer Aussage mit der einer anderen bedeutet. […]* ***Denken heißt nichts anderes als sich eine Gesamtsumme durch Addition von Teilen oder einen Rest durch Subtraktion einer Summe von einer anderen vorstellen.*** *Geschieht dies durch Wörter, so ist es ein Vorstellen dessen, was sich aus den Namen aller Teile für den Namen des Ganzen, oder aus den Namen*