VDI REPORT 17

KÜNSTLICHE INTELLIGENZ

LEITVORSTELLUNGEN UND VERANTWORTBARKEIT

BAND 1: DISKUSSIONSGRUNDLAGE

VDI-HAUPTGRUPPE
DER INGENIEUR IN BERUF UND GESELLSCHAFT

VEREIN DEUTSCHER INGENIEURE VDI

Herausgeber:

Verein Deutscher Ingenieure
VDI-Hauptgruppe
Graf-Recke-Straße 84
40239 Düsseldorf
Tel.: 0211/6214-0

Im Auftrage des VDI herausgegeben von:

Prof. Dr. Armin B. Cremers
Rolf Haberbeck M.A.
Dr.-Ing. Jürgen Seetzen
Prof. Dr. Ipke Wachsmuth

Fachliche Erarbeitung: VDI-Ausschuß "Künstliche Intelligenz"

Redaktion: Dipl.-Pol. Volker M. Brennecke

Diese Veröffentlichung ist kein Objekt des Buchhandels.
Die Abgabe erfolgt zum Selbstkostenpreis von DM 18.--
einschl. gesetzlicher Mehrwertsteuer (Preis für VDI-Mitglieder: DM 15.--)

KÜNSTLICHE INTELLIGENZ

LEITVORSTELLUNGEN UND VERANTWORTBARKEIT

BAND 1: DISKUSSIONSGRUNDLAGE

Inhalt

Einleitung der Herausgeber

„Künstliche Intelligenz" (KI) kann in den nächsten Jahrzehnten in vielen technischen Anwendungsgebieten weitere praktische Bedeutung erlangen. Heute sind nicht nur Fragen nach der gesellschaftlichen Akzeptanz einer entwickelten KI zu stellen, sondern vorerst auch die grundsätzlichen Perspektiven und technischen Realisierungsmöglichkeiten dieses Gebiets zu diskutieren. Da die KI das Ziel verfolgt, die kognitiven Leistungen von Menschen maschinell zu simulieren, ist sie in besonderer Weise auf die anthropologischen Disziplinen (Philosophie, Kognitionspsychologie, Sprachwissenschaft, Neurophysiologie etc.) angewiesen. In diesem wechselseitigen Erkenntnisprozeß besteht die Chance einer Technikgestaltung, die sich später als sozialverträglich erweisen sollte.

Künstliche Intelligenz erfordert, Aspekte menschlicher Intelligenz so zu beschreiben, daß sie durch ein künstliches System simuliert werden können. Bedingt durch diese Ausgangslage, ist das Bild des Menschen in der Künstlichen Intelligenz vor allem ein Bild eines Teiles seiner Intelligenz, wobei die Intelligenzauffassung entscheidend von der Sichtweise eines informationsverarbeitenden Systems geprägt ist.

Künstliche Intelligenz befaßt sich auch mit der Konstruktion von informationsverarbeitenden Systemen, die kognitive Leistungen erbringen, um die theoretisch entwickelten Konzepte und Techniken nutzbringend einzusetzen. Auf informationsverarbeitende Maschinen werden Eigenschaften menschlicher Intelligenz, etwa Schlußfolgerungsfähigkeiten, übertragen, um dadurch geistige Tätigkeiten des Menschen zu unterstützen, zu verstärken oder gar teilweise zu entlasten bzw. zu ersetzen.

Wie bei jeder Technologie ergeben sich Fragen danach, wie der Mensch davon in seinem Selbstverständnis berührt wird. Durch den Anspruch, geistige Tätigkeiten des Menschen zu formalisieren, tritt eine neue Qualität technischer Mittel hervor, wodurch sich auch die Frage nach der Verantwortbarkeit stellt.

Indem nicht nur aus einer ethischen Perspektive, sondern auch aus psychologischer, sprachwissenschaftlicher oder erkenntnistheoretischer Sicht die Möglichkeiten und Grenzen eines Vergleichs der Informationsverarbeitung von Mensch und Maschine erarbeitet werden, können sich Kriterien für die Gestaltung technischer Systeme und der Interaktion dieser Systeme mit dem Menschen ergeben.

Eine grundlegende Diskussion des Themas „Das Menschenbild in der KI", d.h. die Herausarbeitung der expliziten wie auch der vage vorhandenen Vorstellungen der KI-Forschung und Anwendung über den Menschen, soll es ermöglichen, erste Hinweise auf spätere Gestaltungsperspektiven sowie mögliche Auswirkungen zu geben. Könnten nämlich gewisse Annahmen der KI-Forschung vor allem über die Simulationsmöglichkeiten menschlicher Intelligenz in ihren Grenzen besser erkannt werden, ließen sich Folgewirkungen und technische Möglichkeiten ihrer ingenieurmäßigen Umsetzung bereits jetzt besser abschätzen.

Es hat sich in der bisherigen Diskussion gezeigt, daß es *das* Menschenbild in der KI natürlich nicht gibt, ebenso wie es auch in der Anthropologie, in der Wissenschaft vom Menschen, nicht *das* Menschenbild geben kann. Aber es gibt verschiedene Sichtweisen, die sich ergänzen oder widersprechen und die zu betrachten und in das Wechselgespräch, den Diskurs, zu bringen wichtig ist.

Mit dem Ausschuß „Künstliche Intelligenz" versucht die VDI-Hauptgruppe einen Beitrag zu leisten, frühzeitig in einem unübersichtlichen Gebiet den notwendigen Diskurs zwischen den Disziplinen und Institutionen herzustellen. Sowohl von der fachlich interdisziplinären Aufgabenstellung der Hauptgruppe wie dem angestrebten berufspolitischen Ziel, Hilfestellung und Orientierung bei der Verantwortungsübernahme zu geben, handelt es sich bei der Ausschußarbeit um ein genuines Tätigkeitsfeld des VDI im Bereich „Mensch und Technik".

Zentrale Problemkreise der Ausschußarbeit

Bei der Einschätzung der Möglichkeiten und Folgen der Künstlichen Intelligenz setzt sich der Ausschuß mit folgenden zentralen Problemkreisen auseinander:

Erstens: Wo steht der philosophisch-anthropologische Diskurs hinsichtlich der Frage nach dem Menschenbild?

Im Rahmen der Entwicklungen zur „Künstlichen Intelligenz" ist wiederholt die Frage aufgeworfen worden, insbesondere von Winograd/Flores und den Brüdern Dreyfus in Amerika aber auch in Deutschland von Coy, Luft, Krämer, inwieweit das anthropologisch-philosophische Menschenbild, das sich besonders durch die europäische Philosophie nach der Aufklärung und dem Idealismus (Descartes, Leibniz, Kant, Schelling, Fichte, Hegel) und seit Mitte des vorigen Jahrhunderts durch Schopenhauer, Nietzsche, Husserl und in diesem Jahrundert durch Wittgenstein, Heidegger, Gadamer, Jonas sowie in Frankreich durch Foucault, Derrida und Lyotard herausgebildet hat, mit den meist implizit vertretenen rationalistisch-positivistischen Grundvorstellungen der KI-Entwickler über den Menschen in Übereinstimmung oder im Gegensatz steht. Es zeigt sich, daß es gravierende Unterschiede in den Sichtweisen gibt, die sich ganz besonders in der Sprachphilosophie jedoch auch in der Frage nach der Intentionalität des Menschen (Motivation, Wille, Ethik) aufzeigen lassen. Aber auch die Fragen nach dem was „Wissen" ist, oder nach den Grenzen der Erkenntnis werden ganz verschieden gesehen. Dies sind zentrale Probleme, wenn Fähigkeiten von Menschen in einem umfassenderen Sinn technisch simuliert werden sollen, wie Mustererkennung, Lernen, logisches Schließen, Wissensverarbeitung und automatisches Entscheiden.

Es ist sicherlich für die Einschätzung der Möglichkeiten und Grenzen der KI von erheblicher Bedeutung, die Sichtweisen der Anthropologie und Philosophie einschließlich der Wissenschaftstheorie und der Sprachwissenschaft den Grundannahmen der Kognitionswissenschaften und Neurowissenschaften sowie der Künstlichen Intelligenz gegenüberzustellen.

Es dürften sich dabei auch für die Selbsteinschätzung der Ingenieure und der Informatiker, die an diesen Entwicklungen arbeiten, wichtige Einsichten ergeben. Abgesehen davon kann man erwarten, daß prinzipielle oder relative Grenzen dieser Entwicklungen rechtzeitig in den Blick kommen.

Zweitens: Welches „Menschenbild" spiegelt sich in den Voraussetzungen der Bemühungen um die Entwicklung der Künstlichen Intelligenz?

Es könnte sich herausstellen, daß Wunschvorstellungen mancher KI-Protagonisten hinsichtlich der Möglichkeiten der KI irreal sind oder daß das Menschenbild, das sich im anthropologisch-philosophischen Diskurs gebildet hat, angesichts der KI-Entwicklung der Revision bedarf. Beides wäre von grundlegender Bedeutung. Es ist demnach ein intensiver Diskurs zwischen Geisteswissenschaftlern, Naturwissenschaftlern und Technikern erforderlich, um die notwendige Orientierung zu erarbeiten.

Betrachtet man die gegenwärtige KI unter dem Blickwinkel der Fragestellung, inwieweit den unterschiedlichen methodischen Ansätzen und Systemen eine explizite Vorstellung über den Menschen und die Beschaffenheit seines Gehirns und seiner geistigen Vorgänge zugrunde liegt, so finden sich in der Regel allemal diffuse Hinweise auf ein Menschenbild, es sei denn, man betrachtet bereits den überstrapazierten Vergleich des menschlichen Gehirns mit dem Computer als eine entsprechend aussagekräftige Metapher.

In dem klassischen Ansatz der KI, dem Symbolverarbeitungansatz wird von der Hypothese ausgegangen, daß Kognition auf die (regelgeleitete) Manipulation von symbolischen Repräsentationen zurückzuführen sei. Hier wird ein Menschenbild zumindest implizit vertreten, in dem die spezifische biologische Konstitution des Menschen als irrelevant hingestellt und geistige Prozesse als formal zu beschreibende Vorgänge begriffen wurden. Auch die in letzter Zeit zu beobachtende Verlagerung von Interessen von der Ebene der symbolischen Repräsentation hin zur Simulation neuronaler Strukturen, die unter der Bezeichnung „Konnektionismus" bekannt wurde, stellt demgegenüber einzig einen methodologischen, jedoch keinen erkenntnistheoretischen Unterschied dar. Auch hier wird von der gleichen Prämisse einer prinzipiell möglichen Modellierung geistiger Prozesse mit technischen Mitteln ausgegangen.

Die Simulation von Verstandesleistungen auf Computern setzt naheliegenderweise bestimmte Grundannahmen über die jeweiligen mentalen Vorgänge beim

Menschen voraus, die zum einen Teil der individuellen Introspektion der involvierten KI-Forscher, zum anderen Teil der Analyse entsprechender philosophischer, psychologischer, sprachwissenschaftlicher oder neurobiologischer Literatur und Forschung entstammen. Die genaue Betrachtung der dabei vorherrschenden Vorstellungen kann verdeutlichen, daß es keine einheitliche Sicht kognitiver Prozesse gibt. Daraus ergibt sich ein methodischer Zugang für die Betrachtung der oben genannten Fragestellung:

a) die Identifikation der impliziten und expliziten Annahmen der KI über spezifische kognitive Prozesse, z.B. Wissensspeicherung und -verarbeitung, Sprache, Lernen, Problemlösen, sowie die Identifikation des den Annahmen und der Modellbildung zugrundeliegenden Erkenntnisinteresses;

b) die Zusammenführung dieser Einzelergebnisse, um zu einem ganzheitlichen Menschenbild der KI zu gelangen und dessen philosophische, sprachwissenschaftliche und neurobiologische Grundlagen zu bestimmen und damit auch einen Ansatzpunkt für die Lösung anwendungsorientierter Fragen zu gewinnen.

Drittens: Welche Änderungen der Menschenbilder ergeben sich unter dem Einfluß einer "entwickelten" KI?

Solche Fragen betreffen die Gestaltung neuer Technologien der KI und ihre Auswirkung auf menschliche Lebensumstände. Sind sie Konkurrenz für den Menschen, verarmen sie das Bild vom Menschen oder bieten sie gerade die Chance, durch die Orientierung an menschlichen Denkweisen und Weltsichten informationsverarbeitende Systeme zu gestalten, die dem Menschen nicht undurchschaubar und fremd sind?

In diesem Zusammenhang sind folgende exemplarische Fragen zu stellen:

- Wie sieht die Interaktions- und Verantwortungsstruktur aus bei einer neuen Rollenverteilung zwischen Mensch und Maschine, bei der die KI beteiligt ist?

- Ist mit einer Aushöhlung, einem Verlust von Erfahrung und Fachwissen der Menschen am Arbeitsplatz zu rechnen? Wie kann neues Fachwissen heranwachsen, wenn die erforderlichen Erfahrungsfelder durch KI automatisiert sind?

- In welchem Umfang sind die in der KI benutzten Modelle geeignet, Anteile der menschlichen Kreativität und Intelligenz abzubilden?

Diese Schrift wäre ohne das ehrenamtliche Engagement der Mitglieder des VDI-Ausschusses "Künstliche Intelligenz" nicht entstanden. Sie ist vor allem aber nicht eine bloße Beiträgesammlung, sondern der Versuch einer gemeinsam reflektierten Darstellung wichtiger interdisziplinärer Themengebiete der KI. Es ist keine Kompromißschrift geworden, in der alle Meinungsunterschiede nivelliert worden wären, sondern eine produktive Auseinandersetzung, die Aussagen mit hohem Konsens, aber auch mit Dissens darstellt.

An dieser Stelle sei den ehrenamtlichen Mitgliedern des VDI-Ausschusses sowie den Mitarbeitern der Arbeitsgruppen noch einmal sehr herzlich gedankt. Unser Dank gilt auch der VDI-Hauptgruppe - und hier insbesondere Herrn Brennecke - für ihre vielfältigen Unterstützungen. Schließlich wäre das Projekt nicht ohne die finanzielle Unterstützung des BMFT möglich gewesen.

Wir sind an Kritik dieses VDI-Reports sehr interessiert und laden hiermit alle Beteiligten zum Dialog ein. Auf der Grundlage einer Tagung am 10. und 11. September 1992 in Bonn sowie weiterer Diskussionen wollen wir im nächsten Jahr eine verbesserte überarbeitete Buchpublikation vorlegen.

Düsseldorf, im Juni 1992

Die Herausgeber: *A. B. Cremers,*
R. Haberbeck,
J. Seetzen,
I. Wachsmuth

Abschnitt I:

Menschenbild und Computer – Anthropologischer Diskurs

Übersicht zum Abschnitt I

J. Seetzen

Warum wird die Frage nach dem Menschenbild im Blick auf die Computerentwicklung und besonders auf die Entwicklung unter dem Stichwort „Künstliche Intelligenz" gestellt? Ist dies nicht gänzlich praxisfern? Wird der Praktiker nicht zusätzlich durch Ausdrücke wie „Anthropologischer Diskurs" abgeschreckt?

Es mag sein, daß die im ersten Abschnitt vorgelegten Arbeiten für den praktischen Informatiker – trotz aller Bemühung um Verständlichkeit – etwas schwierig zu lesen sind. Aber die verschiedenen wissenschaftlichen „Kulturen" (auch die Informatik) haben nun einmal ihre sprachlichen Eigenheiten.

Informationstechnik allgemein und besonders Computer wirken viel unmittelbarer als andere Techniken in den Bereich menschlichen Verhaltens hinein. Es gibt zu Fragen Anlaß, daß Computer „Geist-Techniken", wie Rechnen, Zeichnen, Entscheiden übernehmen und außerordentlich erweitern, die bis Mitte dieses Jahrhunderts Menschen vorbehalten waren. Am Anfang der modernen europäischen Philosophie sah Descartes in der „denkenden Sache" das Zentrum der menschlichen Existenz. Nun wird dieses Zentrum zum Teil nach außen verlagert.

Es ist das Anliegen dieses ersten Abschnittes, deutlich zu machen, daß es jedoch „Selbst-Verständlichkeiten" wie eine Analogie zwischen menschlichem Denken und Computerprozessen nicht gibt. Je mehr wir über Menschen nicht nur aus der Philosophie und Ethnologie, sondern auch aus der Biologie und Medizin wissen, desto undeutlicher wird das „Selbst-Verständnis" der Menschen. Aber es ist auch nicht zu bezweifeln, daß die Einsichten in die Computertechnik und die Entwicklungen zur Künstlichen Intelligenz zu neuen Sichtweisen für Menschenbilder Anlaß geben.

Die amerikanische Entwicklung zur Künstlichen Intelligenz hat außerordentlich provozierende Äußerungen von Vertretern dieser Entwicklung hervorgebracht. Diese gehen bis zu der Ansicht, wir Menschen könnten froh sein, wenn die entwickelten KI-Produkte uns in einiger Zeit noch als eine Art Haustiere dulden würden. Eine andere besagt, das menschliche Gehirn sei doch schließlich nur eine „Fleisch-Maschine". Als verhältnismäßig selbstverständlich wird in diesen Kreisen angenommen, daß alle intelligenten Fähigkeiten von Menschen im Laufe der Zeit auch von Produkten der KI – und zwar besser als durch Menschen – wahrgenommen werden können.

Es hat in den USA gegen diese Ansichten, die hinsichtlich der Verteilung von öffentlichen Mitteln, besonders aus dem Militäretat der USA, aber auch der zivilen Forschungsförderung in Japan (Fifth Generation Computer Programm), eine nicht unerhebliche Rolle gespielt haben, vernehmlichen Protest seitens einiger Wissenschaftler gegeben. Zum Teil werden diese Stimmen in den folgenden Beiträgen zitiert. Dabei beziehen sich Autoren wie die Brüder Dreyfus oder Winograd und Flores ausdrücklich auf deutsche Philosophen wie Heidegger und Gadamer.

Es erscheint deswegen angemessen, diesen Diskurs in der europäischen Wissenschaftslandschaft aufzugreifen und auf eine breitere anthropologische und philosophische Basis zu stellen.

Menschenbilder sind notwendige Verständigungsversuche über das, was wir von Menschen wissen oder wissen können. Es zeigt sich, daß es verschiedene Perspektiven gibt, aus denen Menschenbilder jeweils anders erscheinen. Es gibt also nicht „das Menschenbild". Eine solche starre Vorstellung wäre mit der Evolution menschlichen Wissens nicht verträglich. Das heißt aber auch, daß die KI-Entwickler ein sehr „einseitiges" Menschenbild haben können. Es geht also in diesem Abschnitt hauptsächlich um die Reibung von Menschenbildern in anderen wissenschaftlichen Domänen und in der Künstlichen Intelligenz.

Das Phänomen des „Geistes" hat in der Philosophie und Wissenschaftstheorie einen bevorzugten Platz, der durch die Entwicklung von Computern in Frage

gestellt wird. Aber bei näherem Hinsehen ist der Computer ein unzureichendes Bild des menschlichen Geistes. Es ist besonders interessant, sowohl in der Geistesgeschichte weiter zurückzuschauen und das Verständnis des Geistes bis in das 17. und 18. Jahrhundert zu verfolgen, weil dort schon die Wurzeln der „Künstlichen Intelligenz" liegen, als auch die moderne Philosophie und Anthropologie, wie sie besonders in Deutschland und Frankreich in diesem Jahrhundert entwickelt worden ist, zu berücksichtigen. Der Fluchtpunkt der anthropologischen Überlegungen liegt in der Möglichkeit verantwortlichen Handelns, also in der ethischen Dimension des Menschseins. Es ist schlechterdings nicht zu erkennen, wie Hervorbringungen der Künstlichen Intelligenz in diese Dimension vordringen sollen.

Auch wenn von Vertretern der sogenannten „harten KI", also denjenigen, die behaupten, daß die KI das volle oder sogar verbesserte Analogon zu Menschen liefern wird, offensichtlich wichtige, relativierende Einsichten aus der Biologie, der Gehirnphysiologie oder der Psychologie außer Acht gelassen werden, so ist doch ein einfacher logischer Beweis, daß dies alles unmöglich sei, nicht zu führen. Dies hat seinen Grund in der Tatsache, daß es der Evolution, wenn auch auf sehr umständlichem und langem Wege gelungen ist, Menschen hervorzubringen. Nur erhebt sich damit die Frage, warum Menschen unbedingt ihr verbessertes Ebenbild schaffen wollen. Sollte die KI nicht bescheidener- und realistischerweise, danach trachten, wie dies für technische Entwicklungen allgemein gilt, die menschlichen Handlungsmöglichkeiten in einem sinnvollen Maße zu erweitern? Welches Maß hierbei allerdings zu beachten ist, ist eine ethische Frage und keine technische. Hierin liegt die eigentliche Herausforderung für den praktischen Informatiker.

Angesichts der Tatsache, daß die reale Künstliche Intelligenz – anders als der Computereinsatz – noch keine besonders spektakulären praktischen Erfolge vorzuweisen hat, ist es an der Zeit, diesen ethischen Diskurs zu führen. Dazu sollen die folgenden Beiträge anregen.

Menschenbilder

R. Capurro und J. Seetzen

Es ist nicht möglich, sich als bewußt lebender Mensch keine Vorstellung von dem, was Menschen sind, zu machen. Indem Menschen sprechen lernen und „ich" sagen, gewinnen sie im Medium der Sprache einen Selbstbezug. Dieser kann mehr oder weniger reflektiert sein. Eine der grundlegenden philosophischen Fragen lautet: „Was ist der Mensch?" Die biblische Tradition hat hierauf eine Antwort gegeben, die immer wieder in dieser Frage durchklingt. „Gott schuf den Menschen nach seinem Bilde", was wir heute als die Befähigung der Menschen zur Mitverantwortung ausdeuten können. Das Nachdenken über das Menschenbild ist auch stets von religiösen Vorstellungen, die in der Sprache und in der Kultur tradiert werden, mitbestimmt.

Die heutige Philosophie hat den Bezug zur Religion, die noch vor zweihundert Jahren selbstverständlich war, weitgehend verloren und handelt von Anthropologie, dem Sprechen über den Menschen, wenn es um die Vorstellungen über das, was Menschen sind, geht. Anthropologie ist aber auch ein Gebiet der realen Wissenschaft. Es handelt sich hierbei um Gebiete der Biologie, der Medizin, der Psychologie, der Ethologie, der Ethnologie, der Soziologie. Alle diese Wissensgebiete tragen wesentlich zu dem bei, was wir von Menschen und vom Menschen wissen können.

Eine vom Anfang der Überlieferungen des Nachdenkens über das, was Menschen sind, bis heute besonders beunruhigende Frage ist, ob für Menschen, die offenbar Selbstbewußtsein haben, damit eine spezifische Differenz zu den übrigen unbelebten und belebten Naturerscheinungen gegeben ist. Diese Frage geht philosophisch in das Problem der Differenz von Leib und Seele oder genauer in die Frage nach der Individualität, besonders nach der Subjektivität der Menschen über und mündet vor allem darin, was wir als Menschen erkennen können, und

was wir tun sollen. Kant hat die Grundfragen der Philosophie auf die vier einfachen Fragen zurückgeführt (Kant 1800):

- Was kann ich wissen?
- Was soll ich tun?
- Was darf ich hoffen?
- Was ist der Mensch?

Kant ist auch der erste Philosoph gewesen, der eine Menschenkunde unter dem Namen Anthropologie schrieb (Kant 1798).

Es hat sich in der europäischen Philosophie seit Descartes eine Tradition herausgebildet, die wesentlich zur Entstehung der Naturwissenschaften beigetragen hat. Diese Tradition hat die erste der vier Kantschen Fragen in den Vordergrund gestellt. Die Erkenntnistheorie hat in unserer Geisteswelt die zweite und dritte Frage, die nach der Ethik und Religion gewissermaßen überwuchert und die Frage nach dem Menschenbild verklingen lassen. Naturwissenschaft und Philosophie haben sich weitgehend voneinander getrennt und berühren sich im wesentlichen nur noch in der philosophischen Wissenschaftstheorie. Dabei bleibt die Frage beunruhigend, was Wissen ist oder wie das Phänomen „Wissen" sich wissenschaftlich und anthropologisch deuten läßt.

Die Selbsteinschätzung oder die Eigenliebe der Menschen hat nach Freud tiefe Kränkungen erfahren (Freud 1916/17). Die erste war die Kopernikanische Wende, daß die Erde nicht der Mittelpunkt der Welt ist. Die zweite war die Darwinsche Erkenntnis, daß Menschen aus dem Tierreich evolutiv hervorgegangen sind. Die dritte war die Freudsche Einsicht, daß Menschen psychologisch nicht Herr im eigenen Hause sind. Gibt es in unserer Zeit eine vierte Kränkung, daß das Wissen, auf dem in unserer Realität soviel aufgebaut ist, ein Kunstprodukt, ein Artefakt wird und sich von den Menschen selbständig macht? Diese Frage läßt die Entwicklung der „Künstlichen Intelligenz" so brisant erscheinen. Denn es wird von KI-Protagonisten ganz naiv davon gesprochen, daß sie für möglich halten, daß menschliches Wissen, ja alle Elemente menschlicher Persönlichkeit, auf KI-Artefakte gewissermaßen wie Dateien umgeladen werden könnten und

dort ein eigenes Sein gewinnen. Dies wäre keine Kränkung der Selbsteinschätzung mehr, dies wäre die evolutive Überwindung der Menschheit (Minsky 1989, Moravec 1988).

Die Frage nach dem Menschenbild stellt sich deswegen in der Perspektive der technischen Utopie von der Künstlichen Intelligenz noch einmal ganz neu. Auch der Praktiker der Informatik, der an diesen Entwicklungen arbeitet, kommt nicht daran vorbei, sich zu fragen oder fragen zu lassen, welchem bewußten oder unbewußten Menschenbild er bei seinem Tun verpflichtet ist. Dabei geht es nicht mehr an, das anthropologische Wissen unserer Zeit einfach beiseite zu lassen. Wenn Physiker und Techniker bisher – wenn auch unberechtigt – glaubten, anthropologischen Fragen ausweichen zu können, so ist dies den Informatikern nicht mehr zu gestatten, denn sie greifen noch unmittelbarer in das menschlich-technische Wirkungsgeflecht ein.

Die anthropologische Besinnung ist heute aber auch durch die Evolutionsbiologie, die Gehirnphysiologie, die Neuroinformatik herausgefordert (Young 1989, Gierer 1985). Weiter sind durch das, was Chaos-Theorie genannt wird, neue Sichtweisen entstanden (Briggs/Peat 1990). Die Frage nach dem Unerkennbaren oder nach der Unentscheidbarkeit wird in unserem Jahrhundert wieder mit vollem wissenschaftlichen Recht gestellt und zwar philosophisch, mathematisch und naturwissenschaftlich. Die Sicherheit, Phänomene nur in physikalistischer, das heißt nur in materiell-energetischer Weise, betrachten zu können, ist dahin (Küppers 1990, Jonas 1981). Wir müssen die Eigenständigkeit der Information als Kategorie der lebenden Strukturen und Systeme anerkennen, die nicht außerhalb der Physis, der Natur sind, aber nicht reduziert werden können auf physikalistische Aspekte (Seetzen 1992).

Nach dem Menschenbild zu fragen, ist verständlicherweise modern geworden. Capurro hat dies in umfassender Weise dargestellt (Capurro 1989, 1990). Aber auch in den Wirtschaftswissenschaften wird gefragt, wie das Menschenbild der Ökonomen beschaffen ist (Bievert/Held 1991).

Menschen sind nicht aus sich selbst – als Individuen – selbstbewußt. Sie sind eingebunden in Sprach- und Kulturräume, zu denen auch die Wissensbereiche gehören. Die außerordentliche Leistung der Naturwissenschaften besteht darin, daß sich in langen, meist schriftlichen Diskursen zwischen vielen Wissenschaftlern über Beobachtungen und Messungen Wissensbereiche gebildet haben, die intersubjektiv von großer Gewißheit sind. Diese Gewißheit hat dazu verführt, von „Naturgesetzen" und deren „Wahrheit" zu sprechen, obgleich sie bestenfalls eine staunenswerte, aber nicht nachweisbar vollständige Isomorphie zwischen menschlichen Modellvorstellungen und realen Erscheinungen sind. Weil Menschen meinen, einen Einblick in die Naturgesetzlichkeit gewonnen zu haben, lag auch immer die Versuchung nahe, das Menschenbild auf den jeweiligen Stand der Wissenschaft zu reduzieren. Der Mensch als Maschine, der Mensch als materiell-energetischer Chemismus. Der Mensch als kybernetisches System und neuerdings der Mensch als Automat, beziehungsweise als Computer (Ebbinghaus 1991). Es muß völlig unbestritten bleiben, daß alle diese Aspekte ihre teilweise Berechtigung und Erklärungskraft haben. Was aber zu bestreiten ist, ist daß mit solchen Sichtweisen Menschen, besonders auch in ihrem sozialen Bezug, hinreichend zu beschreiben sind. Wie Jonas gezeigt hat, ist übrigens die ganze Naturwissenschaft im Wortsinne nicht „denkbar", wenn es nicht kommunizierende Menschen mit Selbstbewußtheit gäbe. Die Seele oder der Geist geht also logisch der Leib- oder Materie-Erfahrung voraus (Jonas 1981).

Einfach gesagt, ist die „Beseeltheit" der Menschen, ihre Fähigkeit, sich bewußt und reflektiv verantwortlich oder unverantwortlich zu verhalten, das heißt in bestimmtem Maße „frei" entscheiden zu können, nicht auf die physikalistischen Aspekte zu reduzieren. Diese Aussage muß nicht auf „Metaphysik" gegründet werden. Die Verantwortungsfähigkeit ist auf Grund der überragenden Komplexität des menschlichen zentralen Nervensystems, das sich in der Evolution herausgebildet hat, mit seinen Wahrnehmungsmöglichkeiten, seinen Vorstellungs- und Denkmöglichkeiten, aber auch mit dem Wollen und Fühlen ein beobachtbares Phänomen, das letzlich auf die „Informiertheit" der Menschen zurückzuführen ist, aber nicht als „Begleitphänomen" physikalistischer Prozesse erklärt werden kann.

Literatur

Bievert B.; Held, M. (Hrsg.) (1991). Das Menschenbild in der ökonomischen Theorie. Frankfurt/Main, New York.

Briggs J.; Peat, F. D. (1990). Die Entdeckung des Chaos. München, Wien.

Capurro R. (1989/90). Menschenbilder. Teil I – III. Mensch-Natur-Gesellschaft 6.II, S. 30; 6.III, S. 61; 7.I, S. 50.

Capurro, R. (1991). Das Menschenbild in den Informationsgesellschaften Ost und West. Baden-Baden: Nomos Verlag

Ebbinghaus K. (1989). Das Menschbild der Künstlichen Intelligenz. Mensch-Natur-Gesellschaft 6.II.

Freud S. (1916/17). Vorlesungen zur Einführung in die Psychoanalyse. In: ders.: Gesammelte Werke. Bd. 11. o.J., S. 294 f.

Gierer A. (1985). Die Physik, das Leben und die Seele. München.

Jonas H. (1981). Macht oder Ohnmacht der Subjektivität Frankfurt am Main.

Kant I. (1798). Der Streit der Fakultäten. Anthropolgie in pragmatischer Hinsicht. In: ders. (1968): Werkausgabe. Hg. von W. Weischedel. Bd. XII. Frankfurt.

Kant I. (1800). Logik. ebd. Bd. VI.

Küppers B.-O. (1990). Der Ursprung biologischer Information. München.

Minsky M. (1989) zitiert nach K. Ebbinghaus (1989).

Moravec H. (1988). Mind Children. London.

Seetzen J. (1992). Information, Kommunikation, Organisation – Anmerkungen zur „Theorie der Informatik". In: W. Coy, et.al. (Hrsg), Sichtweisen der Informatik.

Young J. Z. (1989). Philosophie und Gehirn, Basel, Boston, Berlin.

Der Geist als Computer

A. Kemmerling

Einleitung

Die These, der Geist sei ein Computer, wirft hinsichtlich ihres Inhalts drei gundlegende Fragen auf: was heißt hier „Geist", was heißt hier „Computer" und was heißt hier „sein"? Diese Fragen finden in der derzeitigen Diskussion unterschiedliche Antworten. Die einflußreichste, am heftigsten umstrittene Antwort auf diese drei Fragen besagt folgendes:

(1) Zum *Geist* sind hier jedenfalls alle höheren kognitiven Kapazitäten und deren Ausübung zu rechnen, z.B. das deduktive und induktive Schlußfolgern, das Lernen, das Sprechen, das Problemlösen und dergleichen. Alles andere, was sonst noch unter „Geist" gerechnet werden mag (wie etwa die Bereiche des Fühlens und Wollens), ist nicht unmittelbarer Gegenstand der These.

(2) Unter einem *Computer* ist am besten eine universale Turing- oder von Neumann-Maschine zu verstehen. Solche Maschinen sind Konstrukte von äußerster Schlichtheit und Effektivität. Zumal eine universale Turingmaschine verfügt über nichts als die bescheidensten Grundfertigkeiten der Zeichenmanipulation (einen Strich machen, einen Strich löschen und dergleichen). Sie läßt sich aber so einstellen („programmieren"), daß sie formale Probleme lösen kann – und zwar (im Prinzip) jedes beliebige solche Problem, das überhaupt mit endlichen Mitteln lösbar ist.

(3) Das Wörtchen „sein" ist hier ganz wörtlich zu nehmen. Die These ist nicht, daß der Geist einem Computer gleicht, sondern daß er einer ist.

Diese These, wie wir sie jetzt vorläufig erläutert haben, ist eine spezielle Ausformung der einflußreichsten Theorie über die Natur des menschlichen Geistes, die es in unsern Tagen gibt: des sog. *Funktionalismus*. Der Funktionalismus wiederum läßt sich am besten verstehen, wenn man berücksichtigt, daß er aus dem

Behaviourismus hervorgegangen ist, der von den späten Zwanzigern bis in die Sechziger Jahre unseres Jahrhunderts in der Psychologie und Philosophie von großem Einfluß war.

Der Behaviourismus und seine Engpässe

Der Behaviourismus ist eine ontologisch minimalistische Theorie über die Natur des menschlichen Geistes. Denn alles typisch Geistige (wie z.B. Wünsche, Erwartungen, Entschlüsse, Schmerzen und Farbempfindungen) sollte – gemäß dem Behaviourismus – letztlich nichts anderes sein als die Disposition eines Organismus zu beobachtbarem Verhalten. Zur Veranschaulichung ein simples (und simplifizierendes) Beispiel. Wer den Wunsch hat, ein Glas Bier zu trinken, hat die Disposition,

- ein Glas Bier zu bestellen, wenn er in einer Kneipe ist und genug Geld dabei hat;
- um ein Glas Bier zu bitten, wenn er auf einem Empfang gefragt wird, was er trinken möchte;
- in den Keller gehen, wenn er zuhause ist und nur noch im Keller Bier vermutet;
- und so weiter.

Allgemein gesprochen, er hat die komplexe Disposition, in gewissen Situationen ein gewisses beobachtbares und physikalisch beschreibbares Verhalten an den Tag zu legen. Und der Behaviourismus behauptet noch etwas mehr: der Wunsch, ein Glas Bier zu trinken, ist überhaupt nichts anders als diese komplexe Disposition.

Die Disposition sollte sich letztlich mittels einer Auflistung der physikalisch beschreibbaren Reize (= die verschiedenen Situationen) und der physikalisch beschreibbaren Körperreaktionen (= das jeweilige Verhalten) erfassen lassen. Was am Wunsch nach einem Glas Bier letztlich nur dran ist, was es hier ontologisch betrachtet nur gibt, das sind – laut dem Behaviourismus – die einwirkenden

Reize und die ausgelösten Reaktionen. Dazwischen liegt nichts von eigentlich psychologischem Interesse.

Der Niedergang des Behaviourismus begann (in den Sechziger Jahren) mit der Verbreitung zweier Einsichten. Erstens gibt es keine Aussicht auf eine endlich angebbare Reiz/Reaktion-Liste für die meisten psychischen Phänomene. Zweitens ist jede einzelne Feststellung, die auf einer derartigen Liste aufgeführt werden könnte, durch unabsehbar viele Ausnahmen bedroht (wer ein Glas Bier will, sich in einer Kneipe befindet und genug Geld dabei hat, mag ja dennoch aus vielerlei Gründen kein Bier bestellen – etwa weil seine Sparsamkeit stärker ist als sein Durst, usw.). Wie könnte man diese Ausnahmen theoretisch in den Griff bekommen? Eines war klar: nicht durch die Spezifikation physikalischer Faktoren. Bestenfalls durch die Spezifikation psychischer Faktoren: „Wer ein Glas Bier will, sich in einer Kneipe befindet und genug Geld dabei hat *und wer keine dem entgegenstehenden Wünsche und Überzeugungen hat*, der wird ein Glas Bier bestellen"; das klingt schon viel weniger ausnahmsträchtig. Selbst in einer dispositionalen Theorie des Geistes scheint das Vokabular unverzichtbar.

Ein drittes Problem bestand darin, daß das eigentlich Geistige am Geist im Behaviourismus verloren geht. Denn Dispositionen zu bestimmten physikalisch beschreibbaren Reaktionen unter bestimmten physikalisch beschreibbaren Umständen hat auch ein Virus oder ein Stein. Was kennzeichnet den Unterschied zwischen den genuin geistigen und andern Dispositionen? Dieses dritte Problem war für einen überzeugten Behaviouristen allerdings zweitrangig, denn als Materialist war es ihm mehr darum zu tun, das Geistige als etwas der materiellen Welt Zugehöriges zu erweisen, als darum, den besonderen Charakter des Geistigen herauszustellen.

Vom Behaviourismus zum Eliminativismus

Es gab für den Behaviouristen einen radikalen Schritt, mit dem er all diese Probleme loswerden konnte. Er konnte nämlich einfach leugnen, daß es die geistigen Phänomene, soweit sie seinem dispositionalen Ansatz Schwierigkeiten

bereiten, überhaupt wirklich gibt. Wenn sich der Wunsch nach einem Glas Bier einer behaviouristischen Rekonstruktion prinzipiell widersetzt, so schließt der Behaviourist daraus, daß es solch einen Wunsch – genau besehen – gar nicht gibt. In unserer vorwissenschaftlichen Alltagspsychologoie fingieren wir zwar die Existenz solcher Sachen wie Wünsche, Überzeugungen und Empfindungen; aber wissenschaftlich gesehen gibt es dergleichen genausowenig, wie es Hexen oder Himmelssphären gibt. Aus einer wissenschaftlichen Theorie des Geistes müssen Begriffe wie „Wunsch", „Überzeugung", usw. eliminiert werden, weil sie letztlich nichts bezeichnen.

Diese Auffassung wird *Elimininativismus* (im Hinblick auf geistige Phänomene) genannt. Der Eliminnativismus ist eine prima facie unattraktive Position, weil er unser Bild vom Menschen als einem denkenden, fühlenden und (gelegentlich) planvoll handelnden Wesen völlig preisgibt. Es wird im Eliminativismus unerklärlich, wie es sein kann, daß dieses angeblich grundfalsche Bild bei der Erklärung und Prognose menschlichen Verhaltens über Jahrtausende hin so erfolgreich sein konnte.

Vom Behaviourismus zum Funktionalismus

Der Funktionalismus war ein Vesuch, die eliminativistische Konsequenz des Behaviourismus zu vermeiden, ohne in einen metaphysischen *Körper/Geist-Dualismus* zurückzuverfallen. Der Körper/Geist-Dualismus, wie er seit dem siebzehnten Jahrhundert propagiert worden war, ist die Lehre, daß der Geist etwas Immaterialles und nicht im Raum Vorhandenes sei, das auf irgendeine Weise mit der räumlich gegebenen Welt der Materie in Wechselwirkung oder in Gleichgang sich befinde. Diese Auffassung ist nicht weniger unattraktiv als der Eliminativismus.

Der Funktionalismus steuert einen Kurs zwischen der Skylla des Eliminativismus und der Charybdis des Dualismus. Der Begrifflichkeit des Geistigen wird im Funktionalismus volle theoretische Dignität zugebilligt; die Existenz und echte kausale Wirksamkeit geistiger Phänomene wird nicht in Frage gestellt. Zugleich

wird nicht bestritten, daß jede Ursache und jede Wirkung in der Welt etwas räumlich Lokalisierbares und physikalisch Beschreibbares ist. Dieser Mittelweg wird durch folgende zentrale Unterscheidung eröffnet: der Unterscheidung zwischen der *funktionalen* (oder auch *kausalen*) *Rolle*, die etwas in einem System innehat, und der *materiellen Realisierung* solch einer Rolle. Die funktionale Rolle des Wunsches nach einem Glas Bier ist dadurch charakterisiert, daß sie im Gesamtsystem geistiger Zustände eine spezielle Rolle spielt: wenn dieser Wunsch mit bestimmten weiteren Wünschen und Überzeugungen zusammen auftritt, wird er dazu führen, daß ein Bier bestellt wird; wenn er im Verbund mit andern Wünschen und Überzeugungen vorkommt, wird er zu einem ganz andern Verhalten führen. Dieser Wunsch hat typische Ursachen und Wirkungen, er gehört in ein charakteristisches Kausalgefüge. Und nun die weiterführende funktionalistische Einsicht. Dieses charakteristische Kausalgefüge umfaßt in nicht eliminierbarer Weise andere geistige Phänomene, die selbst wiederum durch ihre Rolle in diesem Kausalgefüge gekennzeichnet sind. Geistiges läßt sich nicht hinwegreduzieren.

Die funktionalistische Lehre besagt: Wird ein geistiges Phänomen in seiner Eigenschaft als geistiges Phänomen thematisiert, so ist von einer funktionalen Rolle die Rede. Wird ein geistiges Phänomen in seiner Eigenschaft als Bestandteil der materiellen Wirklichkeit thematisiert, so ist von einer materiellen Realisierung solch einer Rolle (z.B. einem Zustand des Zentralnervensystems) die Rede. Dieser Unterschied steht in enger Analogie zu dem Unterschied zwischen dem Schachkönig, soweit er durch die Schachregeln gekennzeichnet ist, und einer geschnitzten Figur, die in einer Schachpartie König ist. Der Schachkönig ist durch seine Rolle im System des Schachspiels charakterisiert; „Schachkönig" ist ein Funktionale-Rolle-Begriff. Ein bestimmtes Stück Holz realisiert diese Rolle in einer Partie; „Schachkönig-Figur" ist ein Materielle-Realisierung-Begriff.

Die Schachanalogie weist auf einen weiteren attraktiven Aspekt der funktionalistischen Grundidee hin: dieselbe funktionale Rolle kann auf sehr unterschiedliche Weisen materiell realisiert werden. Wie es für Schachfiguren gleichgültig ist, ob sie aus Holz, Alabaster oder Blei sind, so sind auch geistige Zustände wesentlich flexibel im Hinblick auf die materielle Beschaffenheit ihrer Realisierung.

Eine extraterrestrische Kreatur mit einer ganz anderen chemischen Konstitution als der unseren könnte dennoch einen Geist besitzen.

Vom Funktionalismus zur Computeranalogie

Dem Funktionalismus zufolge ist der Geist wesentlich ein funktionales System. Ein funktionales System überführt in Abhängigkeit vom inneren Ausgangszustand, in dem es sich befindet, Inputs in Outputs. Das System Geist überführt Warhnehmungen in Verhalten; alles, was dabei an psychischen Faktoren wie Stimmungen, Zielen, Vorwissen, Erinnerungen und dergleichen eine Rolle spielt, gehört zu diesem System, das wir „Geist" nennen.

Die funktionalistische Auffassung drängt den Vergleich des Geistes mit einem Computer, der mit einem bestimmten Programm geladen ist, geradezu auf. Denn der programmgeladene Computer läßt ebenfalls diese beiden Betrachtungsweisen zu: erstens die funktionale Betrachtungsweise, wonach er wesentlich dadurch charakterisiert ist, daß er Eingabeketten in bestimmter Weise in Ausgabeketten überführt; und zweitens die physikalische Betrachtungsweise, wonach er ein Gerät mit bestimmten materiellen Eigenschaften (u.a. Ausdehnung, Gewicht, verarbeitetes Silizium usw.) ist.

Der geladene Computer, als funktionales System, ist eine spezielle Turingmaschine. Er kann nur eine beschränkte Anzahl von Sachen. Der mit dem Textverarbeitungsprogramm geladene Computer kann dies und das und jenes nicht, der mit dem Graphikprogramm geladene Computer kann jenes, aber nicht dies und das. Im Gegensatz dazu ist der Computer selbst, wiederum als funktionales System, ein potentieller Alleskönner; er kann potentiell alles, was auf ein Programm paßt. Der Computer selbst ist eine universale Maschine.

Entsprechend läßt sich ein einzelner menschlicher Geist zu jedem einzelnen Zeitpunkt seiner Biographie als eine spezielle Turingmaschine betrachten. Er kann zu jedem einzelnen Zeitpunkt nur eine begrenzte Anzahl von Sachen. Aber auch der Geist, in Abstraktion von seinen einzelnen Zeitstadien betrachtet, ist ein

potentieller Alleskönner. Letztlich kann der Geist alles lernen, was lernbar ist. Unter dieser Annahme, die zwischen Analytizität und Eitelkeit schillert, wird der Vergleich von Geist und Computer noch suggestiver: Der Geist ist eine universale Maschine. (Lernen ist Programmiertwerden oder Sichselbstumprogrammieren.)

Damit ist der Weg vom Behaviourismus zur These vom Geist als Computer in den allergröbsten Zügen nachgezeichnet. Fassen wir kurz und thetisch die hervorstechenden Punkte zusammen:

(1) Der Behaviourismus ist nur um den Preis eines Eliminativismus haltbar.

(2) Der Eliminativismus ist inaktzeptabel.

(3) Der metaphysische Körper/Geist-Dualismus ist inakzeptabel.

(4) Der Funktionalismus eröffnet einen dritten Weg, der

  (a) mit dem Materialismus/Physikalismus verträglich ist;

  (b) eine Unterscheidung zwischen Geistigem und (bloß) Physischem zuläßt;

  (c) die stoffliche Ungebundenheit geistiger Phänomene respektiert;

  (d) eine reichhaltige Analogie von Geist und Computer nahelegt.

Eine reichhaltige Analogie ist immer von theoretischem Interesse, aber nie selbst schon eine interessante Theorie. Die Analogie von Geist und Computer, die der Funktionalismus eröffnet, macht da keine Ausnahme.

Von der Computeranalogie zur These vom Geist als Computer

Es verdient vielleicht, eigens betont zu werden, daß der Funktionalismus den Vergleich des Geistes mit einem Computer zwar nahelegt, selbst aber weder explizit noch implizit die These beinhaltet, der Geist sei ein Computer. Der Funktionalismus enthält ja nicht die Annahme, daß geistige Prozesse Symbolverarbeitungsprozesse sind; gerade diese Annahme ist aber das Herz jeder echten Auffassung vom Geist als Computer.

Eine verbreitete Begründung für diese Annahme sieht in groben Zügen folgendermaßen aus: Ein wesentliches Charakteristikum geistbegabter Lebewesen ist es, flexibel auf den Informationsgehalt dargebotener Reize reagieren zu können. Eine rein physikalische Beschreibung eines Reizes reicht nicht aus, um die nachfolgende Reaktion vorherzusagen oder zu erklären. In einer Vorhersage oder Erklärung der Verhaltensreaktion eines geistbegabten Organismus unterstellen wir, daß er dem Reiz gewisse Informationen entnimmt und diese in sein bereits vorhandenes System von Informationen einordnet, um dann die Verhaltensweise dazu auszusuchen, die für ihn angesichts des *neuen* Informationsstandes im Lichte seiner Präferenzen optimal ist. Mit andern Worten: Wir unterstellen in unserer Erklärung, daß der Organismus Information verarbeitet und daß die in ihm ablaufenden Informationsverarbeitungsprozesse tatsächlich kausalen Einfluß darauf haben, welche Reaktion er zeigt. Diese Prozesse haben also einen entscheidenden Doppelcharakter: sie müssen sensitiv sein für den semantischen Gehalt der involvierten Faktoren (denn unsere Erklärung greift auf den Inhalt der Wahrnehmung, des Hintergrundwissens, der Wünsche usw. zurück) und sie müssen physische Wirkungen (z.B. Körperbewegungen) hervorrufen. Sie haben zugleich einen semantischen und einen physischen Aspekt. Das einzige uns verständliche Modell für ein materielles System mit diesem Doppelcharakter ist der programmierte Computer. (Denn der tut ja nichts anderes – wenn er ordentlich programmiert wurde – als in gesetzmäßiger Weise Symbole auf Grund ihrer physischen Beschaffenheit, aber unter Respektierung ihres Inhalts, zu manipulieren.) Also bleibt uns gar keine Wahl, solange wir nicht auf eine Theorie geistbegabter materieller Systeme verzichten wollen, als die Annahme vom symbolmanipulierenden Geist zumachen.

Geist und Computer: Kognitionswissenschaft und starke KI-These

Dem Funktionalismus zufolge ist der Geist wesentlich ein funktionales System; jede beliebige materielle Realisierung dieses Systems ist ein System mit einem Geist. Müßte dann nicht auch jedes von Menschenhand geschaffene Gerät – wenn es den passenden funktionalen Reichtum besitzt – als etwas gelten, das

einen Geist hat und somit dem Menschen in dieser wesentlichen Hinsicht gleichgestellt ist?

In diesem Zusammenhang ist es sinnvoll, drei Fragen auseinanderzuhalten: (1) Ist es möglich, ein Gerät zu entwickeln, das die gleiche funktionale Organisation aufweist wie ein menschlicher Geist? (2) Hätte solch ein Gerät denn wirklich einen Geist? (3) Ist die funktionalistische Theorie des Geistes auf diese These festgelegt?

Die Antwort auf die erste Frage muß lauten: „Im Prinzip ja. (Ob jedoch eine realistische Aussicht auf derlei besteht, ist eine reine Spekulation.)". Es bedürfte noch ungeheurer Fortschritte auf verschiedenen Feldern, um die funktionale Organisation des menschlichen Geistes auch nur in Umrissen zu erkennen. *Welcher* ingenieurwissenschaftlichen Fortschritte es noch bedürfte, um ein entsprechendes Gerät zu entwickeln, ist überhaupt nicht abzusehen.

Die Antwort auf die zweite Frage ist eine bloße Sache des Glaubens bzw. der metaphysischen Vorlieben. Die sogenannte *starke KI-These* besteht gerade darin, hier mit einem uneingeschränkten Ja zu antworten. Diese These ist also selbst von keinem erkennbaren erfahrungswissenschaftlichen Wert, sondern taugt besser zu Propagandazwecken für KI-Forschungsprogramme.

Die dritte Frage ist so zu beantworten: „Nein, der Funktionalismus ist nicht auf die starke KI-These festgelegt. Aber er ist sehr gut verträglich mit dieser These". Es ist nützlich, an dieser Stelle einen Unterschied zwischen zwei Varianten des Funktionalismus zu machen. Der *methaphysische Funktionalismus* ist die These, daß geistige Phänomene ihrem Wesen nach einzig und allein funktionale Phänomene sind. Der *methodische Funktionalismus* besagt, daß die Betrachtungsweise geistiger Phänomene als funktionaler Phänomene für die Zwecke der psychologischen Forschung angemessen ist. Der methodische Funktionalismus ist der Ausgangspunkt aller Forschungen in den sogenannten Kognitionswissenschaften; und er ist mit einer Ablehnung des metaphysischen Funktionalismus sehr wohl verträglich.

Aber auch der metaphysische Funktionalismus ist nicht auf die Behauptung festgelegt, ein Computergerät, das mit einem hinreichend komplexen Programm geladen ist, habe einen Geist. Ein Anhänger des metaphysischen Funktionalismus könnte allein schon deshalb bestreiten, daß ein Computer einen Geist haben kann, weil ein Computergerät streng genommen weder wahrnehmen, noch handeln kann, daß ihm also schon die charakteristischsten Input- und Output-Funktionen des Geistes abgehen.

Eigentlich müßte man in diesem Zusammenhang eine ganze Reihe sehr verschiedener Thesen klären und erörtern:

(a) *Funktionalistische These*: Jeder Geist ist ein funktionales System von ungeheurer Komplexität.

(b) *Computerthese*: Jeder Geist ist eine universale Symbolverarbeitungsmaschine mit einem ungeheuer komplexen Programm.

(c) *Analogiethese*: Es ist psychologisch aufschlußreich, den menschlichen Geist als eine universale Maschine mit einem ungeheuer komplexen Programm zu betrachten.

(d) *Simulationsthese*: Es ist möglich, eine derart präzise Theorie kognitiver Prozesse zu entwickeln, daß sie sich in Form eines Computerprogramms formulieren läßt.

(e) *Starke KI-These*: Jede universale Maschine mit einem ungeheuer komplexen Programm ist ein Geist.

Es ist bei alldem sehr wichtig, derlei Thesen über den Geist nicht mit neurowissenschaftlichen Hypothesen (oder Analogien) vom *Hirn* als Computer zu verwechseln. Geist und Hirn sind etwa so verschieden wie Blick und Auge, oder anders gesagt: Theoriebildung in der Psychologie und Theoriebildung in der Neurobiologie geschehen auf gänzlich unterschiedlicher Ebene und weitgehend unabhängig voneinander. Es ist sehr wohl denkbar, daß die beste psychologische Theorie uns den Geist (bei der Ausübung seiner kogninitiven Fähigkeiten) als eine Symbolverarbeitungsmaschine darstellt, während das Hirn (bei denselben Gelegenheiten) gemäß der besten neurobiologischen Theorie als ein konnektionistisches Netzwerk aufgefaßt werden muß.

Die Grenzen der Computeranalogie

Die Analogie paßt nur auf einen sehr kleinen Ausschnitt des Geistigen; sie liefert eine anregende Idee, wie man sich die Ausübung gewisser kognitiver Fähigkeiten vorstellen kann. Sollte sie über diesen schmalen Bereich hinaus auf nicht minder wichtige Komponenten des Geistes (wie Emotion, Wille, bildliche Vorstellung usw.) angewandt werden, wäre unklar, worin die Analogie eigentlich bestehen soll.

Und ein Aspekt, der für die Dimension des Geistigen besonders charakteristisch ist, paßt gar nicht zur Computeranalogie: das bewußte Erleben beim Wahrnehmen und Empfinden. Nehmen wir an, jemand trinkt mit geschlossenen Augen aus einem Glas; er soll uns sagen, was er da getrunken hat. Wenn er nun zutreffendermaßen antwortet: „Das ist Limettensaft", dann hat es gewiß Sinn, darüber zu spekulieren, welche Faktoren für seine Antwort eine Rolle spielten und welche Schritte bei der Antwortfindung in welcher Reihenfolge durchlaufen wurden. Und solche Spekulationen kann man in der Art eines Computerprogramms ausformulieren. Soweit mag die Computeranalogie passen. Keinen klaren Sinn hingegen hätte es, wenn behauptet würde, der Computer (oder das Programm) liefere passende Analogien auch dafür, worin das Gechmackserlebnis der betreffenden Person bestand.

Die Betrachtung des Geistes als einem Computer liefert also gewiß keinen Ansatzpunkt für eine einheitliche Theorie des gesamten Phänomens Geist. Zudem gilt es hier eine Reihe sehr unterschiedlicher Behauptungen auseinanderzuhalten, die sich hinter der unklaren These verbergen können, der Geist sei ein Computer.

Literatur

Anderson, A. R. (Hrsg.) (1964). Minds and machines. Englewood Cliffs, N.J.
Boden, M. (1981). Minds and mechanisms. Brighton.
Boden, M. (1987). Artificial intelligence and natural man. 2. Aufl., London.
Boden, M. (Hrsg.) (1990). The philosophy of artificial intelligence. Oxford.

Dennett, D. (1978). Brainstorms. Cambridge, Mass.
Fodor, J. (1987). Psychosemantics. Cambridge, Mass.
Gunderson, K. (1971). Mentality and machines. Garden City,N.Y.
Haugeland, J. (1985). Artificial intelligence: The Very Idea. Cambridge, Mass.
Lewis, D. (1989). Die Identität von Körper und Geist. Frankfurt a.M.
Lycan, W. G. (Hrsg.) (1990). Mind and cognition. Cambridge, Mass.
Mohyeldin-Said, K. A. et al. (Hrsg.) (1990). Modelling the Mind, Oxford.
Penrose, R. (1989). The emperors new mind. Oxford.
Putnam, H. (1988). Representation and reality. Cambridge, Mass.
Pylyshyn, Z. W. (1985). Computation and cognition. 2. Aufl., Cambridge, Mass.
Searle, J. (1986). Geist, Hirn und Wissenschaft. Frankfurt a.M.

Eine weitere kopernikanische Wende?
Zur philosophischen Ortsbestimmung künstlicher Intelligenz

S. Krämer

Künstliche Intelligenz – Technik oder Wissenschaft?

Die künstliche Intelligenz vereinigt heterogene Forschungsbemühungen. Einerseits ist sie eine ingenieurwissenschaftliche Disziplin, beschäftigt mit dem Design symbolischer und maschineller Systeme, die dazu dienen, geistige Tätigkeiten des Menschen zu automatisieren. In dieser Perspektive produziert Künstliche Intelligenz „Denkzeuge", sie liefert Geistestechnologien.

Andererseits versteht sie sich als eine theoretische Disziplin, sie stellt Theorien und Modelle auf über das, was „Geist", „Denken" und „Intelligenz" sei. Mit diesem von ihr konzipierten Bild vom Geist praktiziert die Künstliche Intelligenz ein Stück Metaphysik.

Künstliche Intelligenz zwischen Geistestechnologie und Metaphysik – dieses Spannungsverhältnis begegnet uns wieder, sobald wir versuchen, den Ort der Künstlichen Intelligenz im philosophischen Diskurs der Neuzeit zu bestimmen.

Interpretiert als eine Geistestechnologie knüpfen Ideen der Künstlichen Intelligenz geradezu nahtlos an Elemente der Erkenntnistheorie der philosophischen Aufklärung im 17. Jahrhundert an (Krämer 1988). Interpretiert als Metaphysik des Geistes, welche die Idee verabschiedet, daß Geist gebunden sei an ein Selbstbewußtsein, an die Personalität und Subjektivität desjenigen, der Geist zeigt, reiht die Künstliche Intelligenz sich ein in den subjektkritischen Diskurs der Moderne, den wir mit den Namen Nietzsche, Freud, Heidegger und mit den französischen postmodernen und dekonstruierenden Theoretikern verbinden. Ein Diskurs, in welchem die für die Philosophie der Aufklärung so typische Ver-

knüpfung von Geist und Subjekt als der „Sündenfall" neuzeitlicher Geistesgeschichte desavouiert wird (Frank 1986).

Dieser Stellungswechsel Künstlicher Intelligenz im philosophischen Disput tritt exemplarisch hervor in seinem Verhältnis zur Philosophie des Rationalismus. Gemeinhin werden die Väter des Rationalismus, Descartes und Leibniz, zu den Vorreitern der Idee Künstlicher Intelligenz erklärt. Doch die Sachlage ist komplizierter. Denn die Künstliche Intelligenz steht sowohl in der Erbfolge des Rationalismus, wie sie auch als dessen Subversion gelten kann – abhängig alleine davon, ob wir die Künstliche Intelligenz in praktisch-technischer Absicht als eine Geistestechnologie oder in theoretisch-wissenschaftlicher Absicht als Modell und Metaphysik des Geistes interpretieren.

Künstliche Intelligenz in der Erbfolge der rationalistischen Erkenntnistheorie

Schon immer ist das Denken, soweit es sich an der wissenschaftsförmigen Rationalität orientiert, angewiesen auf das künstliche Medium der Schrift (Goody et al. 1986). Die Schrift aber ist – anders als die „natürliche" Sprache – ein Artefakt; und das Schreiben wird zu einer Technologie. Die alphabetische Schrift dient als eine Technologie der Kommunikation: Durch den geschriebenen Text wird das Wissen abgetrennt vom Wissenden und über die Distanzen des Raumes und der Zeit kommunizierbar. Anders die operative Schrift, zu der alle formalsprachlichen Zeichensysteme, zum Beispiel das dezimale Positionssystem oder die Buchstabenalgebra gehören. Hier dient die Schrift nicht alleine dazu, die natürlichen Grenzen der Kommunikation zu erweitern. Vielmehr sind operative Schriften Werkzeuge des Denkens; sie sind eine Geistestechnologie. Gefordert ist dazu allein, daß eine bestimmte Domäne von Gegenständen isomorph den Zeichen eines Kalküls zugeordnet werden kann, so daß die gesetzmäßigen Zusammenhänge zwischen diesen Gegenständen explizierbar werden in Gestalt der Formations- und Transformationsregeln des Kalküls. Wo dies der Fall ist, können Probleme gelöst werden durch algorithmisches Abarbeiten von Kalkülre-

geln. Mentale Tätigkeiten können als handgreiflich-technischer Prozeß der Symbolmanipulation bewerkstelligt werden.

Operative Schriften, die beim Problemlösen zum Einsatz kommen, haben die Funktion symbolischer Maschinen. Jedes durch eine symbolische Maschine operationalisierbare Verfahren ist auch durch eine wirkliche Maschine realisierbar. Die Evolution der Geistestechnik symbolischer Maschinen ist die Vorgeschichte der Künstlichen Intelligenz (Krämer 1991a).

Mit dem schriftlichen Rechnen im dezimalen Positionssystem im 14. und 15. Jahrhundert, wird der Gebrauch einer symbolischen Maschine zu einer Kulturtechnik. Im 16. und 17. Jahrhundert finden formale Problemlösungsverfahren Eingang in nahezu alle Bereiche der exakten Wissenschaften. Der Kalkül wird zum Signum des wissenschaftlichen Geistes der Neuzeit (Krämer 1991b).

Descartes und Leibniz sind zwei Pioniere der mathematischen und logischen Kalkülisierung im 17. Jahrhundert. So wundert es nicht, daß diese rationalistischen Philosophen die mechanischen Verfahren einer symbolischen Maschine auch in nichtlogischen und nicht-mathematischen Kontexten des Denkens fruchtbar zu machen suchten.

Descartes konzipiert eine mathesis universalis, eine universalwissenschaftliche Disziplin, deren – quantifizierbare – Gegenstände nur noch gegeben sind in einer von ihm eigens entworfenen künstlichen Sprache und deren Verfahren konzipiert sind als Operationen innerhalb dieses kunstsprachlichen Mediums (Descartes, Regeln). Erkenntnisse könnten so mit einer Sicherheit gewonnen werden, wie sie sonst nur den Garantieprodukten algorithmischer Herstellungsverfahren eigen sind. Allerdings ist Descartes' Kunstsprache eine an der Geometrie orientierte Bildersprache, die Größenverhältnisse analog repräsentiert. Dadurch soll beim Operieren mit den Symbolen die Aufmerksamkeit festgehalten werden für das, was die Symbole jeweils bedeuten.

Leibniz teilt mit Descartes die für die rationalistischen Erkenntnisprogramme typische Intention, Wahrheit auf Richtigkeit zurückzuführen. Doch anders als für

Descartes, liegt für Leibniz die Pointe formaler Symboloperationen gerade darin, daß die Aufmerksamkeit des Geistes für das, was die Symbole bedeuten, absentiert werden könne, so daß wir uns im Kalkül dem Ariadnefaden mechanischer Symbolmanipulationen blind überlassen, uns beim Vollzug kalkülisierter Verfahren wie Maschinen verhalten können (Leibniz, Charakteristik). Mit den Worten von Leibniz: Wo wir beim Denken Kalküle einsetzen, werden wir zu „spirituellen Automaten" (Leibniz, Metaphysik) . Solche operativen Dienste leisten Kalküle allerdings nur, sofern diese nicht nach geometrischen, vielmehr nach algebraischem Vorbild aufgebaut werden, als ein konsequent digitales Repräsentationssystem.

Mit seinem calculus ratiocinator entwirft Leibniz einen Universalkalkül des wissenschaftlichen Denkens, der es ermöglichen soll, daß alle wahren Sätze automatisch ableitbar werden, sowie von jedem vorliegenden Satz entscheidbar werde, ob er wahr oder falsch sei (Hermes 1969).

Die mit der mathematisch-logischen Grundlagendiskussion unseres Jahrhunderts nachgewiesene Unmöglichkeit solch universaler Denkmaschine steht hier nicht zur Diskussion. Für uns ist allein von Bedeutung, in welcher Weise der Idee Künstlicher Intelligenz philosophisch vorgearbeitet wird. Deren Credo, daß die menschliche Wissensverarbeitung vollständig explizierbar sei, nimmt Leibniz in den folgenden Annahmen vorweg:

(1) Mit der operativen Schrift löst sich das Wissen nicht nur ab vom Wissenden; vielmehr kann Wissen überprüft und neues Wissen gewonnen werden mit der Geistestechnik eines operativen symbolischen Systems. Voraussetzung dessen ist allerdings die Propositionalität des Wissens.

(2) Wo dieses System als Kalkül, somit als Maschine organisiert ist, können semantik-orientierte Verfahren reduziert werden auf rein syntaktische Symbolmanipulationen.

(3) Wo immer ein Denkprozeß vollständig formalisierbar ist, kann er auch von einer wirklichen Maschine realisiert werden. Formalisierung und Mechanisierung sind Begriffe gleicher Extension.

Kein Zweifel: Mit der rationalistischen Idee, das schlußfolgernde Denken in einem formalen und prinzipiell mechanisierbaren symbolischen System auszuführen, ist eine Erkenntnistechnik anvisiert, die systematisch verwandt ist mit der Künstlichen Intelligenz. Doch diese verwandtschaftlichen Bande beschränken sich ausschließlich auf die Technik des Erkennens, deren Pointe (nach Leibniz) darin liegt, ohne Geist exekutierbar zu sein. Sie greifen gerade nicht über auf den Begriff des Geistes selbst. Obschon für Descartes und Leibniz gewisse kognitive Leistungen organisiert werden sollen nach dem Vorbild maschineller Operationen, wird die Maschine ihnen gerade nicht zur Metapher oder zum Modell für das, was Geist ist.

Künstliche Intelligenz als Subversion des Rationalismus

Das Vermögen, Geist zu haben, wird in der rationalistischen Philosophie an die Eigenschaft gekoppelt, ein Subjekt zu sein. Dabei impliziert „ein Subjekt zu sein" für die Philosophie der Aufklärung zweierlei: 1. Das Vermögen „ich" zu sagen und fähig zu sein zu einem Selbstbewußtsein, 2. das Vermögen verantwortlicher Urheber von Handlungen zu sein und damit den Status einer Person zu haben.

Was das heißt, sei beispielhaft illustriert an Descartes' Position. In seinem methodischen Zweifel wirft Descartes alle für Irrtum überhaupt anfälligen, und das heißt korrigierbaren Urteile über Bord (Decartes, Meditationen). Als letzter und einziger Rettungsanker dafür, daß es so etwas wie „Wissen" überhaupt geben kann, bleibt ihm das „cogito", das „ich denke" übrig. Denn auch der radikalste Zweifel bleibt – und dies ist unbezweifelbar – ein Fall des Denkens. Erst durch die unmittelbare Gewißheit, die im Vollzug dieses Aktes von Bewußtwerdung meiner selbst als Denkendem liegt, ist ein Fundament geschaffen, auf welchem dann das Gebäude des positiven Wissens haltbar errichtet werden kann.

Worauf es hier ankommt, ist der Sonderstatus des Selbstbewußtseins als eine besondere Art von Wissen. In seiner Unkorrigierbarkeit, in seiner Privilegierung der Ich-Perspektive, ist das Selbstbewußtsein gar keine Form eines propositiona-

len Wissens (Frank 1991, Shoemaker 1984). Wobei „Propositionalität" impliziert, daß das Wissen a) wahr oder falsch sein kann und b) durch Beschreibungen in der Perspektive der dritten Person singular vollständig darstellbar ist. So zeigt Descartes in seinem Rekurs auf die nicht eliminierbare Subjektivität des Denkens die Grenzen allen propositionalen Wissens auf und damit die Grenzen der Annahme von der vollständigen Explizierbarkeit des Wissens.

Doch auch die zweite Bedeutung von „Subjektivität" im Sinne von „handelnd verantwortlich sein" spielt eine Rolle bei Descartes. Descartes unterscheidet zwei Vermögen des Geistes: Verstand und Wille (Descartes, Meditationen). Zum Verstand zählt alles, was mit der Tätigkeit des „intellectus" zu tun hat: Empfinden, Einbildung und reines Denken. Zum Willen gehört alles das, was zusammenhängt mit der Tätigkeit der „voluntas": Begehren und Ablehnen, Behaupten und Bezweifeln. Beim Erkennen arbeiten Verstand und Wille zusammen; der Verstand liefert Ideen, der Wille aber bestätigt oder verwirft sie, verwandelt also die Ideen in Urteile. So kann Irrtum entstehen, wo die Domäne des Willens beginnt; nämlich dann, wenn der Wille mit seiner Zustimmung den beschränkten Bereich der klaren und deutlichen Ideen überschreitet.

Folge dieser Urteilstheorie ist eine Analogisierung von Irrtum und Sünde, die Descartes tatsächlich durchführt. Der Irrtum wird zur philosophischen Form einer Schuld, deren theologische Version die Sünde ist. Fehler des Denkens erhalten den Status moralischer Verfehlungen.

Erinnern wir uns: Descartes geht es um die Zurückführung von Wahrheit auf Richtigkeit. In der Perspektive seiner Erkenntnistechnik hieß dies: Ein vorgegebenes Regelwerk muß beim wissenschaftlichen Denken möglichst konform eingehalten und abgearbeitet werden. Doch in der Perspektive seiner Theorie des Geistes enthüllt sich der nichtreduktionistische, nichtformale Sinn seiner Verknüpfung von Wahrheit mit Richtigkeit: Das Wahre erweist sich als theoretische Spielart und das Gute als die praktische Spielart eines universalen Strebens nach dem Richtigen, dessen „Sitz" das moralisch verantwortliche Subjekt ist. Die Bezugnahme auf eine der Verantwortung fähigen und zur Verantwortung pflichtigen Person wird zum konstitutiven Element des cartesischen Begriffes von Geist.

Daß Selbstbewußtsein, daß die Fähigkeit „ich" zu sagen, eine notwendige Bedingung für Geistigkeit sei, wird von Leibniz in seiner Metaphysik der Monade fortgebildet (Leibniz, Metaphysik). Für Leibniz ist jeder geistige Akt vollziehbar nur vom Standpunkt einer individuellen Perspektive: Individualität wird zu einer unhintergehbaren Voraussetzung für den Besitz von Geist.

Hier ist nicht der Ort, um die im Rationalismus entwickelte philosophische Konzeption des Geistes bis ins Einzelne zu verfolgen. Doch soviel ist sicherlich deutlich geworden: „Geist" gilt den rationalistischen Philosophen als eine anthropozentrische Kategorie. Und „anthropozentrisch" heißt dabei: Bei der Beschreibung von „Geist" spielen Eigenschaften des Subjektes, welches über Geist verfügt, eine nicht eliminierbare Rolle. Das für die Naturwissenschaft konstitutive reduktionistische Erkenntnismodell, welches darauf hinausläuft, den Subjektbezug aller Erfahrung zu eliminieren, endet für die rationalistischen Philosophen, sobald es ihnen um die Erklärung von „Geist" zu tun ist.

Eine kopernikanische Wende?

Die Idee der künstlichen Intelligenz geht von der Annahme aus, daß subjektive Eigenschaften gerade keine notwendigen Bedingungen für den Geist darstellen. Nun lassen sich innerhalb der Künstlichen Intelligenz drei Richtungen ausmachen: die orthodoxe (McCarthy 1979, Newell, Simon 1976), die revisionistische (Minsky 1990, Waltz 1988) und die emergente KI (Rumelhart, McClelland 1986). Diese leisten einen je anders gearteten Beitrag zur Überwindung des subjektorientierten Geistbegriffes. Er läßt sich – kursorisch – so bestimmen, daß das Bild vom Geist als einem autonomen Agenten sukzessive ersetzt wird durch das Bild einer dezentralisierten Agentur von Geistern.

Mit allen dem Modell von „Denken als Informationsverarbeitung" verpflichteten Disziplinen teilt also die Künstliche Intelligenz das Bestreben, an die Stelle eines anthropomorphen einen speziesinvarianten Begriff von Geist zu etablieren. Geistigkeit soll nicht länger ein Definiens des Menschen bleiben, sondern – im Prinzip – auch einer Maschine zugesprochen werden. So scheint die Künstliche

Intelligenz zu einer Wende beizutragen, bei welcher der anthropozentrische Geistbegriff der Aufklärung ersetzt wird durch einen nicht anthropozentrischen. Damit wird – nicht anders, denn bei den Entdeckungen von Kopernikus, Darwin und Freud – der Mensch einer ihm liebgewordenen Vorrangstellung beraubt, er wird desillusioniert in Bezug auf seine Stellung im Kosmos, auf seine Auszeichnung gegenüber anderen Geschöpfen.

Tatsächlich scheinen die Befürchtungen und Ängste, die die Künstliche Intelligenz freisetzt, in eben diesem „kopernikanischen Flair" zu wurzeln. Denn als Geistestechnik, die faktisch zum Einsatz kommt und gesellschaftlich wirksam wird, ist die Künstliche Intelligenz geradezu marginal. Brisant wird sie alleine als eine Metaphysik des Geistes.

. . . und ihr Preis

Diese Brisanz wurzelt nicht bloß darin, daß das Selbstbild des Menschen nachhaltig erschüttert wird. Vielmehr geht es hier um weitreichende ethische Probleme. Die Aufklärung entwickelt einen nicht-intellektualistischen Geistbegriff: Zum Geist gehörten für sie nicht nur Vernunft und Verstand, also Intellekt, sondern auch der Wille, also das, was den Menschen zu einem der Rechenschaft fähigen und zur Verantwortung pflichtigen Wesen macht. Der Kern der Fähigkeit zu geistigem Tun bestand für die rationalistische Philosophie in der Fähigkeit zu moralischem Handeln. Kentnisse könnten nur erworben werden, wo die Fähigkeit zu Erkenntnis, also zu Urteilsbildung gegeben ist. Wissen könne nur existieren, soweit ein Gewissen vorhanden ist.

Wo aber der Geist gleichgesetzt wird mit intelligentem Tun, intelligentes Tun aber spezifiziert wird als Informationsverarbeitung, welche ihrerseits gewonnen wird am Vorbild des Computers oder der Boltzmann-Maschine, werden kognitive Kompetenzen unabhängig von moralischen Kompetenzen projektiert. Die Abtrennung der intellektuellen von den moralischen Fähigkeiten ist der Preis des nichtanthropozentrischen Geistbegriffes. In einer Welt, deren Probleme nicht durch bloß effizientere Informationsverarbeitung zu lösen sind, vielmehr mutige

Entscheidungen für oder gegen Wertsetzungen verlangen, ist dieser Preis zu hoch.

Das intellektualistische, auf operative Informationsverarbeitung reduzierte Bild vom Geist, das in der Künstlichen Intelligenz favorisiert wird, führt hinaus auf eine moralische Amputation an unserem Selbstbildnis.

Literatur

Descartes, R. (1972). Meditationen über die Grundlagen der Philosophie mit sämtlichen Einwänden und Erwiderungen. Hamburg: Felix Meiner Verlag.

Descartes, R. (1979). Regeln zur Ausrichtung der Erkenntniskraft. Hamburg: Felix Meiner Verlag.

Frank, M. (1986). Die Unhintergehbarkeit von Individualität. Frankfurt: Suhrkamp.

Frank, M. (1991). Selbstbewußtsein und Selbsterkenntnis. Essays zur analytischen Philosophie der Subjektivität. Stuttgart: Philipp Reclam jun.

Goody, J.; Watt, I.; Gough, K. (1986). Entstehung und Folgen der Schriftkultur. Frankfurt: Suhrkamp.

Hermes, H. (1969). Idee von Leibniz zur Grundlagenforschung. Die ars inveniendi und die ars iudicandi. Studia Leibnitiana Suppl.,vol. III, S. 92-102.

Krämer, S. (1988). Symbolische Maschinen. Die Idee der Formalisierung in geschichtlichem Abriß. Darmstadt: Wissenschaftliche Buchgesellschaft.

Krämer, S. (1991a). Denken als Rechenprozedur. Zur Genese eines kognitionswissenschaftlichen Paradigmas. Kognitionswissenschaft 2, S.1-10.

Krämer, S. (1991b). Berechenbare Vernunft. Kalkül und Rationalismus im 17. Jahrhundert. Berlin, New York: de Gruyter Verlag.

Leibniz, G. W. (1965). Kleine Schriften zur Metaphysik. Hg. u. übers. v. H.H. Holz, Frankfurt: Insel -Verlag.

Leibniz, G. W. (1966). Zur allgemeinen Charakteristik. In: ders. Hauptschriften zur Grundlegung der Philosophie, Bd.I. Hamburg: Felix Meiner Verlag.

McCarthy, J. (1979). Ascribing mental qualities to machines. In: M. Ringle (ed.) Philosophical perspectives in artificial intelligence. Brighton, Sussex: Harvester Press, S. 161-194.

Newell, A.; Simon, H. (1976). Computer Science as empirical inquiry: Symbols and search. Communications of the Association for Computing Machinery, 19.

Minsky, M. (1990). Mentopolis. Stuttgart: Klett-Cotta.

Rumelhart, D. E.; McClelland, J. L. (1986) Parallel distributed processing: Explorations in the microstructure of cognition. Bd. I u.II, Cambridge (Mass.): MIT Press.

Shoemaker, S. (1984). Identity, cause and mind. Philosophical essays. Cambridge (Mass.): Cambridge University Press.

Waltz, D. L. (1988). The prospects for buildung truly intelligent machines. In: S. R.Graubard (ed.) The artificial intelligence debate. False starts, real foundations. Cambridge (Mass.): MIT Press, S. 191-211.

Künstliche Intelligenz: Eine Phänomenologische Kritik

B. Becker und Ch. Lischka

Einleitung

Betrachtet man gegenwärtige Bemühungen in der Künstlichen-Intelligenz-Forschung näher, so wird deutlich, daß sie hinsichtlich ihrer expliziten oder impliziten Zielsetzung zu einer Denktradition[1] gehören, die nach kompromißloser Ein-Ordnung und Verallgemeinerung von Vieldeutigem und Mannigfaltigem strebt. Dieser Anspruch äußert sich innerhalb der KI in verschiedenen Bereichen und wird zum Beispiel deutlich in der Expertensystem-Entwicklung mit dem dort praktizierten Versuch, rationale Modelle menschlicher Expertenpraxis zu entwickeln und auf diese Weise allgemeine Standards an die Stelle individuellen Verhaltens zu setzen. Das Bestreben, Irrationalitäten in Problemlösungen durch formale Regelungen zu eliminieren – ein expliziter Anspruch der Expertensystementwickler[2] – kann als materialisierter Ausdruck eines Bestrebens interpretiert werden, der auf Überwindung von Widerständigem und Ungeregeltem zielt und an die Stelle des „Etre sauvage" (siehe hierzu Wadenfels 1987) eine klare Ordnung setzen will. Die KI fügt sich damit in eine historische Entwicklung ein, in der durch die Entwicklung spezifischer Ausgrenzungssysteme versucht wurde, „Ungebändigtes" und „Regelloses" (Waldenfels 1987) verfügbar zu machen und Unerwünschtes durch die Institutionalisierung bestimmter – nicht zuletzt machterhaltender – Diskursformen zu eliminieren.

Im Folgenden wird zunächst versucht, an die von Dreyfus, Winograd/Flores und anderen formulierte Kritik der KI anzuknüpfen, indem die klassische und auch

[1] Eine ausführliche Kritik hierzu findet sich bei Foucault (1977).

[2] Hierzu existiert eine Fülle an Literatur, siehe beispielsweise Clancey/Shortliffe (1984), Schwartz/Griffin (1986), Retti/Trost (1987), Puppe (1991).

für die KI aufzeigbare Subjekt-Objekt-Trennung durch Kontrastierung mit dem Merleau-Ponty'schen Leibkonzept in Frage gestellt wird. Anschließend wird die von Waldenfels und Foucault artikulierte Kritik an spezifischen, dem Paradigma der Ordnung unterworfenen Ausgrenzungspraktiken aufgenommen und für eine kritische Einschätzung der KI fruchtbar gemacht. Dabei soll der Nachweis geliefert werden, daß die KI als maschinell unterstützter Versuch der Re-Etablierung einer allgemeinen Ordnung gedeutet werden kann, um anschließend zu zeigen, daß jeglicher Versuch der Errichtung formaler Ordnungsstrukturen vom gelebten Äquivalent menschlicher Praxis immer wieder unterlaufen werden kann.

Künstliche Intelligenz in den Fallstricken des klassischen Subjekt-Objekt-Antagonismus

An anderer Stelle (siehe beispielsweise Winograd, Flores 1986) wurde bereits ausführlicher dargelegt, daß die KI mit ihren Konzepten in den klassischen Antagonismus von Subjekt und Objekt gerät, das heißt eine Trennung von erkennendem Subjekt und objektiv gegebener Welt vornimmt. Diese auf die neuzeitliche Philosophie[3] zurückgehende Subjekt-Objekt-Spaltung verdeutlicht sich in dem für die KI zentralen Begriff der Repräsentation. Subjekte bilden dieser Vorstellung zufolge mentale Repräsentationen über die Welt, die wiederum die Basis für kognitive Operationen darstellen. Entsprechend wird insbesondere im klassischen Ansatz der KI, dem sogenannten Symbolverarbeitungs-Paradigma, Kognition mit der – formalen Regeln gehorchenden – Manipulation mit syntaktisch strukturierten Repräsentationen gleichgesetzt. Damit steht nicht nur ein erkennendes, aktives Subjekt einer passiven Welt gegenüber, sondern dieses aktive Subjekt wird zudem zum letzten Fundament aller Sinnstiftung, zur ordnenden Instanz. Die für die mentalistische Philosophie charakteristische Zentrierung des Subjekts wird von der KI, insbesondere in ihrer kognitionswissenschaftlichen Ausprägung, zumindest implizit übernommen.

Dieser klassischen Subjekt-Objekt-Spaltung setzt die Phänomenologie spätestens seit Heidegger und Merleau-Ponty eine andere Position entgegen, mit der

[3] Insbesondere Descartes; vergl. die entsprechende Kritik Heideggers in Sein und Zeit.

sie einerseits den szientistischen Naturalismus und andererseits das transzendentalphilosophische Erbe[4] beziehungsweise die kritizistische Bewußtseinsphilosophie Kants zu überwinden trachtet. Die Grundthese von Merleau-Ponty lautet: Die Welt erschließt sich dem „Subjekt" durch dessen existentiell bedeutsame Ausgerichtetheit (Intentionalität) zur Welt und durch sein Handeln in der Welt. Erst auf der Basis einer präreflexiven (oder „stummen") Erfahrung und der sich daraus entwickelnden Praktognosie entstehen Repräsentationen – sie sind, ebenso wie die bewußte Reflexion, zumindest in einer onto- oder phylogenetischen Betrachtungsweise sekundär gegenüber dieser primären Welterschließung. Damit kommt aber weder dem Subjekt noch den Dingen eine primäre sinnbildende Funktion zu; vielmehr entsteht Sinn erst im spezifischen Miteinander von „Subjekt" und Welt, oder besser gesagt: in der spezifischen Reversibilität, durch die sich das wechselseitige Übergreifen von Natur und Geist stets neu ausdrückt. „Das Subjekt ist damit nicht mehr zentrale Instanz, sondern selber Moment eines Subjekt und Objekt umgreifenden Sinngeschehens" (Waldenfels 1980).

Die hier sehr abstrakt anmutende Position soll im Folgenden in der notwendigen Kürze näher verdeutlicht werden. Wie oben bereits angedeutet, tritt Merleau-Ponty explizit mit dem Ziel an, die beiden klassischen Alternativen „Existenz als Ding" und „Existenz als Bewußtsein" zu umgehen und eine Vermittlung von objektiv Gegebenem und subjektiv Erlebtem zu erreichen. Konkreter Anknüpfungspunkt ist dabei für Merleau-Ponty der Leib, der weder als Objekt in der Welt noch als reines Bewußtsein gedacht werden kann. In dem von ihm geprägten Begriff der inkarnierten Existenz drückt sich aus, wo für ihn der eigentliche Grund liegt: „Das leibliche Da bedeutet eine Vorgegebenheit von Welt, Selbst und Anderen, hinter die wir nicht zurück können, und fernerhin ist diese Vorgegebenheit kein bloßes factum brutum, gegen das unsere Sinnentwürfe anrennen, vielmehr heben die Prozesse der Sinnbildung selber mit einer leiblichen Spontaneität an und schlagen sich in leiblichen Gewohnheiten" nieder (ebda, S. 17).

Der Leib tritt dabei zunächst als eigener Leib, als „corps propre" auf, der Erfahrungen macht und in der Erfahrung gegenwärtig ist. Er ist aber immer auch ein

[4] Was sich trotz aller gegenteiligen Versuche bei Husserl nach wie vor identifizieren läßt.

„corps objectiv", d.h. ein physikalischer Körper. Damit ist der Leib gleichzeitig Medium zur Welt wie auch Instanz der Verankerung in der Welt. Wenn unser Erleben im Verhalten zur Darstellung kommt, fungiert der Leib einerseits als „sichtbarer Ausdruck eines konkreten Ego" (Merleau-Ponty 1966), die individuelle wie kulturelle Vorgeschichte gleichermaßen verkörpernd, andererseits weist er aber auch äußerlich beobachtete Verhaltensweisen auf. Innen und Außen, sinnhafte Intention und körperliche Mechanismen zeigen sich damit im leiblichen Verhalten integriert. Das zur-Welt-hin-Ausgerichtet-Sein ist nicht allein ein Produkt mentaler Sinnkonstitution, sondern ist in den Strukturen leiblichen Verhaltens selbst mit angelegt – die Einheit von Körper und Geist ist über den Leib vermittelt.

Darüberhinausgehend impliziert die ursprüngliche Intentionalität des Leibes stets eine mögliche Bereitstellung für zielgerichtetes Handeln – der Leib ist auf die Welt hin ausgerichtet. Die Welt läßt sich nicht dinghaft, sondern stets nur situationsbezogen erfahren. Unser Verhältnis zu den Dingen ist somit durch die Art bestimmt, wie wir uns ihnen nähern beziehungsweise die Dinge sich uns nähern. Jegliche Bestimmung resultiert aus der Verankerung des aktiven Leibes einem Gegenstand, einer Situation gegenüber. Daraus folgt: Wir sind nicht in der Welt oder denken – ihr entgegentretend – über die Welt nach; wir sind zur Welt. Sinn erschließt sich dem Subjekt weder rein empirisch noch transzendental intellektualisiert, sondern vermittels seines Leibes, für den das Zur-Welt-Sein sinnkonstitutiv ist. Unser Leib heftet sich der Welt an, ist in sie eingebettet. Diese Einbettung, dieses Einwohnen ist gebunden an die Bewegungs- und Handlungserfahrung unseres Leibes. Unser Zugang zur Welt ist der einer „Praktognosie", welche Merleau-Ponty nicht nur als eigenständig, sondern sogar als ursprünglich begreift[5]. Diese leibgebundene Erschließung der Welt stellt somit das Fundament jeglicher Sinnkonstitution dar – die Dezentrierung des Subjekts ist demnach schon in seiner Leiblichkeit angelegt.

[5] An diesem Punkt weist er eine große Nähe zu Piaget auf, siehe auch Merleau-Ponty (1976).

Die Welt ist also weder als eine „objektiv Gegebene" noch als eine reine „Vorstellungswelt" zu begreifen, sie ist vielmehr in der Struktur meines Leibes schon vorgezeichnet. Selbst wenn in der empirischen Beobachtung oder in der Reflexion über die Welt eine Loslösung von der ursprünglichen Erfahrung erfolgt, so müssen wir doch, um die Welt uns vorstellen oder über sie nachdenken zu können, zuallererst durch unseren Leib in sie eingeführt sein.

Damit ist aber die klassische Position der KI in Frage gestellt: Repräsentationen, die in der KI als strukturierte mentale Entitäten begriffen werden, erscheinen im Blickwinkel Merleau-Ponty's allemal als Resultate einer ursprünglich gedachten Praktognosie, einer durch das leibgebundene Handeln konstituierten primären Welterfahrung. Reflexion ist dieser ursprünglichen Erfahrung gegenüber sekundär – ihr eigentlicher Sinn erschließt sich erst aus der leibgebundenen Welterschließung. Diese Vorgängigkeit leiblichen In-der-Welt-Seins kann beim Versuch, kognitive Prozesse zu beschreiben, zu simulieren oder gar erklären zu wollen, nicht außer Acht gelassen werden.

Eine ähnliche Kritik wird an anderer Stelle, vor allem bei Dreyfus (1985) und – zumindest ansatzweise – bei Winograd/Flores (1986) unter Bezugnahme auf Heidegger artikuliert. Sie führte in den letzten beiden Jahren zur Entwicklung einer methodologischen Alternative, die unter den Etiketten „KI ohne Repräsentation", „pragmatic turn" bzw. „Artificial Life" (vgl. z.B. Artificial Intelligence 47, 1991, Special Volume: Foundations of Artificial Intelligence, S. 1–3) gehandelt wird. Doch stoßen – unabhängig von ihrer unzweifelhaften heuristischen Funktion – derartige Ansätze dort auf Barrieren, wo die von Merleau-Ponty ins Spiel gebrachte vorgängige Leibgebundenheit von Kognition letztes Fundament menschlicher Sinnauslegung bleibt und die Welterschließung die Fähigkeit zum Handeln in der Welt und in Bezug auf Andere voraussetzt. Zudem scheint die zentrale Rolle des Subjekts als sinnkonstitutierende Instanz nach wie vor ungebrochen, da sich die spezifische Ambiguität von Subjekt und Welt, ihr wechselseitiges Aufeinandereinwirken, in diesen alternativen Konzepten kaum wiederfindet. Gerade dieses dialektische Verhältnis durchbricht jedoch erst die klassische Sicht, derzufolge jegliche Form der Sinnauslegung beim Subjekt blieb

und diesem nicht nur sinnbildende, sondern darüberhinaus auch ordnende und reglementierende Funktion bei jeglicher Form der Weltauslegung zukam.

Im Folgenden wird der Versuch unternommen, KI als Ausdruck einer Materialisierung dieser ordnenden Funktion des Subjekts zu interpretieren und aufzuzeigen, daß jeglicher Versuch, eine allgemeine Ordnung zu schaffen, immer wieder durch das besondere Verhältnis von Welt und Subjekt durchbrochen wird.

Künstliche Intelligenz als Ausdruck und Materialisierung einer allgemeinen Ordnungs- und Ausgrenzungspraxis

Betrachtet man nun im zweiten Schritt KI-Konzepte hinsichtlich der ihnen immanenten Zielsetzung etwas genauer, so wird deutlich, daß die KI in gewisser Hinsicht die Funktion einer Ordnungsinstanz übernimmt beziehungsweise übernehmen soll. Regellose Momente zu eliminieren, irrational anmutende Handlungsabläufe in eine geregelte Struktur einzuordnen, Kommunikationsprozesse durch eine strenge Systematisierung von Sprachpraxis festzulegen – dies deutet sich dementsprechend als impliziter Anspruch für die Entwicklung und den Einsatz von KI-Systemen immer wieder an[6] . KI wird damit Ausdruck und Instrument einer Totalität, die jegliche Praxis durchdringt und maßgeblich prägt. Der Versuch, Ungeregeltes einem abstrakten, durch Machtansprüche und Ausgrenzungspraktiken untermauerten Ordnungsprinzip zu unterwerfen, ist nicht neu; wie Foucault, Derrida und Waldenfels in ihren jeweiligen Untersuchungen zeigen, charakterisieren derartige Bemühungen seit langem die menschliche Kulturgeschichte. Zudem läßt sich, unter Bezugnahme auf Foucault's Analyse (Foucault 1977), insbesondere die angewandte KI als Materialisierung einer spezifischen Ausgrenzungspraxis interpretieren, vor allem jener Form der Ausschließung, die sich als Folge einer Teilung in wahre und falsche Diskurse ergibt. Die Unterwerfung sprechender und handelnder Subjekte unter einen festgelegten, als wahr etikettierten Diskurs läßt sich in einem historischen Rückblick immer wieder nachweisen. Dabei stellen die Festlegung von Wahrheitsbedingungen und die Bestimmung von Kriterien für ihre Überprüfung ebenso Selektions-

[6] Ein gutes Beispiel bieten hierfür die Expertensysteme.

momente dar wie das Bemühen, Fremdes, in einen Diskurs nicht unmittelbar Einzuordnendes, als falsch zu klassifizieren und damit letztlich zu eliminieren. In eine derartige Ausgrenzungspraxis ist die KI unschwer einzuordnen. Betrachtet man in concreto Systementwicklungsprozesse, findet genau dies statt: Nicht unmittelbar Verwertbares wird unterdrückt, vermeintlich rationale Diskursformen gegenüber nicht unmittelbar transparenten präferiert und letztlich nur das als gültig betrachtet, was in das technische Modell der Entscheidungsfindung paßt (Becker et al. 1991).

Dieser implizite Ordnungsanspruch und die damit einhergehende, konkret beobachtbare Ausgrenzungspraxis der KI sind in mehrfacher Hinsicht problematisch: Nicht nur hinsichtlich der Perpetuierung traditioneller Ausgrenzungspraktiken und der Formulierung entsprechender Ordnungsansprüche; sondern überdies offenbart sich bei näherer Betrachtung, daß rationale Modelle menschlicher Praxis (und als solche lassen sich die meisten KI-Systeme begreifen) stets ein grundlegendes Defizit durch die Verkennung fundamentaler Eigenarten menschlichen Denkens und Tuns aufweisen. So zeigt insbesondere Waldenfels auf, daß ein Blick auf die menschliche Praxis immer wieder deutlich macht, wie durch den konkreten Sprechakt, die situationsgebundene Handlung und andere Formen individueller Seinsäußerung jeder formalen Regelung ein lebendiges Äquivalent entgegengesetzt wird und auf diese Weise das „Regellose" sich gegenüber dem „Geregelten" zu behaupten vermag.

Bei näherer Betrachtung erweist sich nämlich, daß die menschliche Praxis durch ein nicht zu vernachlässigendes Maß an Ungeregeltem gekennzeichnet ist. Trotz immer wieder unternommener Bemühungen, rationale Standards für Kommunikationsprozesse und Entscheidungsverläufe zu entwickeln beziehungsweise abstrakte Regeln für die Sprach-Praxis zu formulieren, schimmert in concreto stets ein Moment von Ir-Rationalität durch. Zudem sind Ordnungen niemals völlig starr. Sie weisen an ihren Grenzen Spielräume für individuelles Verhalten auf und zeigen zumindest hinsichtlich der Art, wie man sich in ihnen bewegt, ein Potential von Variabilität auf.

Die entgegen allen Ordnungs- und Ausgrenzungsbemühungen unabänderlich vorfindbare Regellosigkeit von Handlung und Kommunikation ist durch die Tatsache mitbedingt, daß jeder formalen Struktur ein konkreter Akt, ein „Gelebtes Äquivalent" (Waldenfels in Grathoff 1976, S. 26) entgegenstehen muß, um überhaupt zu gewährleisten, daß Handlungen Effekte zeigen beziehungsweise Kommunikation und Sprache funktioniert. Unabhängig von allen formalen Codierungsbemühungen und maschinellen Regelungen existiert für jeden Einzelnen immer ein Horizont an Verhaltens- und Ausdrucksmöglichkeiten, ein Freiraum also, der die individuellen Schattierungen hervorruft, durch welche sich eine lebendige Praxis auszeichnet (Waldenfels 1990). Starre, festgelegte Strukturen ersticken die für Kultur lebenswichtige Spontaneität ihrer Mitglieder – ein deterministisches Regelsystem würde damit zum Verlust des notwendigen Verhaltensspielraums einzelner und letztlich zur Erstarrung von Diskursen und Praktiken führen.

Überprüfen wir diesen Aspekt detaillierter: Sicherlich ist nicht zu leugnen, daß Sprache, Handlung und Kommunikation immer auch durch ritualisierte Formen und abstrakt gesetzte Regelungen charakterisiert sind. Jeder individuelle Sprechakt, jedes subjektgebundene Verhalten birgt stets einen Aspekt des Gewohnheitsmäßigen in sich. In jeder Ausdrucksform menschlicher Praxis, in jedem Diskurs lassen sich Wiederholungen feststellen und formale Ordnungsschemata bestimmen, innerhalb derer konkretes Verhalten stattfinden muß (siehe hierzu ausführlicher Foucault 1974). Und dennoch: Jede Handlung birgt auch immer ein originäres, kreatives Moment in sich, das auf den prinzipiell offenen Horizont hindeutet, auf den hin jede menschliche Ausdrucksform ausgerichtet ist (siehe hierzu ausführlicher Waldenfels 1987). Selbst wenn Handlungen in vertrauter oder abgegriffener Manier vollzogen werden, selbst wenn immer wieder stereotypes Verhalten und durch Wiederholungen gekennzeichnete Sprechakte auftreten, so ist doch die Anwendung jeder noch so starr formulierten Regel in Ansätzen schöpferisch, weil stets, wenn auch nur nuanciert, ein individueller Akzent im konkreten Verhalten sichtbar wird. Die intendierte Standardisierung von Sprache und Verhalten, offiziell dokumentiert in sogenannten „Firmenphilosophien", diziplinären Sprachcodes, starren Diskursformen spezifischer kultureller Bereiche wie Wissenschaft, Kunst und Politik, vor allem aber in einem zur

Norm gesetzten Habitus bestimmter gesellschaftlicher Gruppen (vgl. hierzu als Beispiel Bourdieu 1981), ist niemals eine vollständige – jeder sich an diesen gesetzten Standards orientierende Akt umfasst – zumindest potentiell – eine Überschreitung der gesetzten Grenzen. Der individuelle Sprechakt, das konkrete Verhalten produzieren einen Überschuß, der nicht eindeutig rückführbar ist auf die gesetzte Norm. Vor allem in der dialogischen Situation (Waldenfels 1987) existiert jenseits der konkreten Bezugnahme auf den Gesprächspartner im Frage-Antwort-Modus ein Potential ungenutzter Möglichkeiten, sich darin äußernd, daß jede gestellte Frage mehrere mögliche Antworten – neben der konkret gegebenen – zur Folge haben könnte. Dieses Faktum nicht ausgeschöpfter Potentiale gilt auch für jede gesetzte Ordnung. Noch so starr anmutende Ordnungen sind an ihren Grenzen überschreitbar, bergen in der Art, wie ihnen begegnet wird, Variabilität in sich. Damit aber wird Praxis, der individuelle Akt im Besonderen, zum notwendigen Konterpart jeder formalen Regelung.

Dies läßt sich auch auf die KI übertragen. Wenn sie als Versuch gedeutet wird, mit den jeweiligen Systemen eine formale Regelung für Praxisbereiche institutionalisieren zu wollen, muß ihr genau dieses lebendige Äquivalent entgegengesetzt werden. KI im Besonderen wie auch Technik im Allgemeinen darf nicht als Eindringen einer allumfassenden Totalität in die Lebenswelt fatalistisch hingenommen oder mit apokalyptischen Visionen beschrieben werden, sondern bedarf eines Umdenkens hinsichtlich der Art, wie mit Technik umzugehen ist. Was für das konkrete Handeln in institutionalisierten Ordnungen und abgesicherten Diskursen gilt, muß auch für die KI in Anschlag gebracht werden: Technik bietet die Möglichkeit, variable Umgangsformen zu entwickeln, die dazu führen müssen, daß sie weder übermächtige Ordnungsinstanz noch reines Mittel zum – wie immer auch gearteten – Zweck wird. Waldenfels und Dreyfus vertreten hier – teilweise unter Bezugnahme auf Heidegger – ein Gegenkonzept, die „Poiesis" und bezeichnen damit die Erweiterung der Technik hin zu Sozial-, Human- und Körpertechniken beziehungsweise das Zusammenwachsen von Kunst und Technik unter dem leitenden Gedanken der Möglichkeit von Umbildungen gängiger und dem Entstehen neuer, variabler Ordnungen

Hier liegt ein Anknüpfungspunkt für all jene, welche jenseits der fragwürdigen Alternativen einer reinen Instrumentalisierung von Technik und apokalyptischer Visionen einen sinnvollen Umgang mit Technik anvisieren beziehungsweise für möglich erachten.

Resümee

1. Die philosophische Position Merleau-Ponty's macht deutlich, daß die repräsentations-orientierte Sichtweise der KI verkürzt ist, da sie die Leibgebundenheit jeder Kognition außer Acht läßt und dabei verkennt, daß das leibliche In-der-Welt-sein des Menschen und seine intentionale Ausgerichtetheit zur Welt eine wesentliche Bedingung menschlicher Erkenntnis ist. Kognition erscheint unter diesem phänomenologischen Blickwinkel stets mit einer ursprünglichen Handlungserfahrung verknüpft, über die sich Menschen ihre Welt erschließen. Leiblichkeit als Voraussetzung jeder Form des Zur-Welt-Seins kann deswegen bei der Betrachtung kognitiver Akte nicht einfach außer Betracht bleiben, wie dies in klassischen KI-Ansätzen geschieht.

2. KI-Systeme als technische Produkte führen eine besondere Form der Ausgrenzung fort, die in der rationalistischen Denktradition schon lange angelegt ist. Dies wird beispielsweise bei der Entwicklung von wissensbasierten Systemen deutlich, wo all die Aspekte menschlichen Wissens außer Acht gelassen werden, die sich nicht in formale Strukturen integrieren lassen. Das Besondere, das Individuelle, aber auch das Ungeregelte, das in der menschlichen Praxis neben allem Geregelten auch immer existiert, wird so zugunsten des Allgemeinen, des Geordneten vernachlässigt oder eliminiert.

3. Diese Kritik hat nicht nur Folgen für die Methodik der KI, besonders für die Kognitionswissenschaft, sie muß auch hinsichtlich der Art, wie mit den entsprechenden technischen Produkten umgegangen wird, zum Umdenken führen. In der Kognitionsforschung deutet sich bereits mit dem erwähnten „pragmatic turn" eine Änderung an. Welche Gestaltungs- und Nutzungsalternativen sich hieraus für den technischen Bereich ergeben, muß sich erst noch zeigen.

Literatur

Becker, B.; Steven, E.; Strohbach, S.(1991). Leitvorstellungen in der Wissensakquisition. Diagnose und Kritik. Wissmod Projektbericht Nr. 3, St. Augustin.

Bourdieu, P. (1981). Homo Academicus. Frankfurt.

Clancey, W.; Shortliffe, E. (eds.)(1984). Readings in medical artificial intelligence. The first decade. Reading, Mass.: Addison-Wesley.

Dreyfus, H. (1985). Was Computer nicht können. Königstein.

Foucault, M. (1974). Die Ordnung der Dinge. Frankfurt.

Foucault, M. (1977). Die Ordnung des Diskurses. Frankfurt.

Grathoff, R. H. (1976). Maurice Merleau-Ponty und das Problem der Struktur in den Sozialwissenschaften. Stuttgart.

Merleau-Ponty, M. (1966). Phänomenologie der Wahrnehmung. Berlin.

Merleau-Ponty, M. (1976). Die Struktur des Verhaltens. Berlin.

Puppe, F. (1991). Einführung in Expertensysteme. 2. Aufl. Berlin usf.: Springer.

Retti, J.; Trost, H. (Hg.)(1987). 3. Östereichische Artificial-Intelligence-Tagung (Proceedings). Berlin usf.: Springer.

Schwartz, S.; Griffin, T. (1986). Medical thinking. The psychology of medical judgement and decision making. Berlin, etc.: Springer.

Waldenfels, B. (1980). Spielraum des Verhaltens. Frankfurt.

Waldenfels, B. (1987). Ordnung im Zwielicht. Frankfurt.

Waldenfels, B. (1990). Der Stachel des Fremden. Frankfurt.

Winograd, T. ; Flores, F. (1986). Understanding computers and cognition. New Jersey.

Ethischer Ausblick

J. Seetzen und R. Capurro

Kant hat die Philosophie eine Wissenschaft des Menschen, seines Vorstellens, Denkens und Handelns genannt. Handeln heißt auf griechisch „praxis". Philosophie ist also ursprünglich immer auch ‚praktische Philosophie'. Praxis hat mit Gewöhnung, mit Sitten zu tun. Ethik, vom griechischen Wort ethos ‚Gewöhnung' hergeleitet, ist also der Bereich der Philosophie, in dem Menschen als handelnde Wesen betrachtet werden (Capurro 1990).

Es hat sich eingebürgert, philosophisch zwischen Moral und Ethik zu unterscheiden, obwohl ‚mores' im Lateinischen nichts anderes als Sitten heißt. Aber die zweite Kantsche Frage „Was soll ich tun?", die im Abschnitt ‚Menschenbilder' erwähnt worden ist, hat nicht immer eine Antwort in der Gewohnheit des Verhaltens und in den tradierten Sitten und Wertvorstellungen.

Es gibt häufig erstmalige Situationen, in denen nicht auf bekannte Verhaltensmuster zurückgegriffen werden kann, sondern diese geradezu verheerend wirken würden. Hier muß ethische Reflexion einsetzen, das heißt das ‚vorsichtige' Nachdenken, was die Folgen des Handelns für andere und den Handelnden selbst sein können. Der berühmte Kantsche kategorische Imperativ ist eine Formel für reflektiertes Handeln, nämlich nur nach derjenigen Maxime zu handeln, von der man wollen kann, daß sie zum allgemeinen Gesetz werde.

Reflexion kann im Sinne der Überlegungen des Anthropologen Gehlen (1966) als inneres Probeverhalten, meist als inneres Sprechen, verstanden werden. Das heißt, in einer neuen Situation ‚besprechen' wir uns mit uns selbst. Platon hat überhaupt das Denken als ‚das Sprechen der Seele mit sich selbst' (Platon 1988, Bd. 6, S. 119) bezeichnet. Reflexion setzt uns auch erst in den Stand, über die Beweggründe unseres Handelns Antwort geben zu können, also verantwortlich

zu handeln. Das setzt aber voraus, daß auch unser Denken und unsere Sprache in die „Verantwortung" genommen werden müssen (Capurro 1988).

Bei der ethischen Reflexion können wir uns nur zum Teil auf Wissen stützen, und keinesfalls können unsere naturwissenschaftlichen Modellvorstellungen diese Reflexion leiten, sondern wir sind darauf angewiesen, das Wollen und Fühlen anderer Menschen und mehr und mehr auch die Folgen für unsere Mitgeschöpfe und Lebensbedingungen zu bedenken. Naturwissenschaftler und Techniker sind geneigt, der Rationalität ihrer Disziplin auch in ethischen Fragen den Vorrang zu geben. Sie übersehen dabei in der Regel, daß ethische Probleme – das, was wir tun sollen – nicht durch das Erkennen, sondern durch das Wollen und Getriebensein bestimmt sind.

Nun kommen wir heute mehr als früher gerade auch durch die systemhaft wirkenden Informationstechniken in die mißliche Situation, daß wir als Einzelmenschen die Folgen unseres Handelns nicht überschauen können. Dies gilt sicher ganz besonders für Hervorbringungen der „Künstlichen Intelligenz". Jonas hat in seiner Schrift ‚Das Prinzip Verantwortung' (Jonas 1984) auf diesen Sachverhalt hingewiesen und gefordert, die räumlichen und zeitlichen Fernwirkungen unseres Handeln in die ‚Vorsicht' mit aufzunehmen.

Dies ist leichter gesagt als getan. Hardware und Software als Produkte der Informatik sowie ihre systemhafte Verknüpfung durch die Telekommunikation haben Anwendungsmöglichkeiten und Wirkungen, die der Entwickler und Hersteller unmöglich in seinen Fernwirkungen überschauen kann, weil die Anwendungen immer auch durch die Zielsetzungen und besonderen Verhältnisse der Anwender beeinflußt werden. Ob mit einem Textverarbeitungssystem Poesie oder ein Geschäftsbericht oder Anleitungen zu einem Terrorakt oder einfach Unsinn produziert wird, kann der Entwickler nicht beeinflussen. Ob eine Mustererkennung auf der Basis von KI in der Produktion, in der Medizin, beim Militär oder als Überwachung von Menschen eingesetzt wird, hat der Entwickler entsprechender Software nicht zu vertreten.

Wenn wir auf der kollektiven Ebene nicht in weitgehende Verantwortungslosigkeit geraten wollen, müssen die Kollektive – Organisationen, Unternehmen, Verwaltungen, Parteien, Regierungen – zur Verantwortung gerufen werden. Dies ist ohne Frage ein Machtspiel, das nur funktioniert, wenn den Mächten, die sich kollektiv bilden, entsprechende Gegenmächte gegenüberstehen. Von Wissenschaftlern und Technikern, aber auch von Wirtschaftlern und anderen Betroffenen, ist deswegen zu verlangen, daß sie sich auf entsprechenden Ebenen in Organisationen, die zur Verantwortung rufen können, mit ihrem Wissen, aber auch mit ihrer Vorstellung von wünschbarer Zukunft engagieren. Der Informatiker, der zum Beispiel KI entwickelt, ist nicht einfach Privatmann, der sich der kollektiven Ethik entziehen darf. Er muß sein Wissen in den kollektiven ethischen Diskurs einbringen.

Dem Postulat nach einer kollektiven Ethik, bei der das Subjekt der Verantwortung auf Gruppen, Organisationen, Staaten übergeht, muß also durch einen kollektiven, diskursiven Reflexionsprozeß, durch denkendes und ausgesprochenes Probeverhalten auf kollektiver Ebene entsprochen werden. Was unter dem Schlagwort der Technikfolgenabschätzung verhandelt wird, ist – noch wenig bewußt – ein Ansatz zur kollektiven Reflexion. Wenn wir unser immer wirkungsvoller technisch instrumentiertes Handeln noch verantworten wollen, dann muß die kollektive ‚Vorsicht', das diskursive Bedenken des Folgenden und damit das ‚Zur-Verantwortung-Rufen' zu einer Selbstverständlichkeit werden. Davon sind wir leider zum Schaden unserer Zukunft noch weit entfernt.

Literatur

Capurro, R. (1988). Die Verantwortung des Denkens. Forum für interdisz. Forschung 1, S. 15-21.

Capurro, R. (1990). Ethik und Informatik. Univ. Stuttgart, Antrittsvorlesung.

Gehlen A. (1966). Der Mensch. Frankfurt a.M., Bonn.

Jonas H. (1984). Das Prinzip Verantwortung.

Platon (1988). Sophistes. In: ders.: Sämtliche Dialoge. Hrsg. v. O. Apelt. Bd. 6. Hamburg.

Abschnitt II:

Menschenbilder in der real existierenden KI

Real existierende KI – Menschenbilder, Leitvorstellungen, Konzeptquellen
Übersicht zum Abschnitt II

R. Haberbeck

Unter dem Begriff „real existierende Künstliche Intelligenz" werden alle Ansätze zusammengefaßt, die sich als Künstliche Intelligenz verstehen oder unter Definitionen von Künstlicher Intelligenz fallen, die im Zeitraum von 1956 bis 1992 vorzufinden sind, sowie detaillierte Planungen, die bis zum Jahr 2000 reichen. Dieses Konzept will nicht eine Klärung oder Diskussion der vielfältig verstandenen Definitionen der Begriffe Intelligenz oder Künstliche Intelligenz anstreben. Das Ziel der Betrachtung und Einschätzung der Technologiefolgenabschätzung und Verantwortbarkeit der KI erfordert eine dermaßen weite und offene Herangehensweise, um nicht von vornherein Aspekte in dieser relativ neuen Fragestellung auszuschließen, die auf den ersten Blick vielleicht unwichtig erscheinen, aber sich als bedeutend herausstellen können. Weiterhin ist kein ausgearbeitetes Konzept von KI vorzufinden, das allgemein anerkannt und unumstritten ist.

Die Betrachtung der real existierenden KI ist eingebettet in den anthropologisch-philosophischen Diskurs des Menschenbilds der KI, der einen Rahmen für die Konzeptquellen, Leitvorstellungen und Menschenbilder der real existierenden KI bildet. Seetzen, Capurro/Seetzen, Kemmerling, Krämer und Becker/Lischka behandeln diese Aspekte im Abschnitt I dieses Bandes. Die Zukunftsauswirkungen der Künstlichen Intelligenz, die über den Bereich der real existierenden KI hinausgehen und den Zeitraum bis zum Jahre 2030/2050 ins Auge fassen, werden von Wachsmuth, Wilker, Kremeier, Görz, Röpke, Schreiber und Strube in Abschnitt III thematisiert. Diese Untersuchungen überschneiden sich teilweise mit dem Gegenstand der Betrachtung der real existierenden KI. Inwieweit die real existierende KI das Schicksal des real existierenden Sozialismus teilen wird – die Ausprägungen abstrakter Leitvorstellungen und Menschenbilder werden

von der Dynamik der Markt- und Anwendungsorientierung absorbiert – wird die Zukunft und somit die Geschichte zeigen.

Der Begriff Künstliche Intelligenz kann in diesem Zusammenhang ohne eine präzise Definition des Intelligenzbegriffs verstanden werden als eine Orientierung der maschinellen Informationsverarbeitung am Menschen. Diese Formulierung unterscheidet Künstliche Intelligenz von der traditionellen maschinellen Informationsverarbeitung, die sich ausschließlich an der einfachen technischen Machbarkeit oder abstrakt logisch-mathematischen Prinzipien orientiert (z.B. Orientierung an Formalismen oder Automatentheorien). Dieser weite Ansatz soll alle Ausprägungen der KI umfassen.

Die vorzufindenden Ansätze der KI in diesem weiten Verständnis können nach ihren zugrundeliegenden Konzeptquellen und Leitvorstellungen eingeteilt werden. Jeder Intelligenzbegriff und jeder Begriff von Künstlicher Intelligenz beinhaltet ein Menschenbild. Diese Menschenbilder sind durch Leitvorstellungen und Konzeptquellen bestimmt, die ihre Vertreter nicht immer in vollem Umfang erfassen und die in vielen Aspekten nicht vollständig transparent sind. Somit wird in diesem Zusammenhang ein dynamisches und historisches Konzept von Künstlicher Intelligenz verwendet, das der Entwicklung und Vielfältigkeit des Gegenstandes Rechnung trägt. Die Betrachtung der Leitvorstellungen, Menschenbilder und Konzeptquellen der Künstlichen Intelligenz in einem weiten Sinn soll eine Basis schaffen für die Einschätzung der Leitvorstellungen und der Verantwortbarkeit von KI.

Künstliche Intelligenz ist in diesem Sinne nicht eindeutig als Forschung, Technik oder Anwendung in Bezug auf maschinelle Informationsverarbeitung zu verstehen sondern hat einen die Disziplinen übergreifenden Charakter. Im Gegensatz zu interdisziplinären Ansätzen wie z.B. der Bio-Chemie, in der die naturwissenschaftliche Erkenntnis des Zusammenhangs von Biologie (insb. Genetik) und Chemie die Grundlage für die Schaffung des interdisziplinären Gebiets Bio-Chemie war, ist die Künstliche Intelligenz als ein transdisziplinärer Ansatz zu verstehen. Künstliche Intelligenz basiert als Ansatz nicht auf gesicherten wissenschaftlichen Ergebnissen zwischen bisher isolierten wissenschaftlichen

Disziplinen, die diese Disziplinen unter einem gemeinsamen, systematischen Aspekt darstellen (interdisziplinärer Ansatz). Künstliche Intelligenz im weiten Sinne verstand und versteht sich als Anspruch, zwischen verschiedenen wissenschaftlichen Disziplinen, den entsprechenden technischen Realisierungen und den Anwendungen verbindende und übergreifende Konzepte und Lösungen zu erreichen. In diesem Sinne geht KI über den interdisziplinären Ansatz hinaus, weil erstens eine Verbindung nicht nur zwischen *wissenschaftlichen* Disziplinen thematisiert ist, und weil zweitens nicht eine abgesicherte, wissenschaftliche Erkenntnis die Grundlage des Überschreitens von Grenzen von Disziplinen ist, sondern ein programmatischer Anspruch: Die Orientierung der maschinellen Informationsverarbeitung am Menschen.

Folgende Konzeptquellen erscheinen für die Erfassung der Ausprägungen der KI unter Berücksichtigung des transdisziplinären Charakters von besonderer Bedeutung.

1) Die formal-mathematische Logik als Konzeptquelle

Diese Konzeptquelle teilt die Künstliche Intelligenz in grundlegenden Aspekten mit der traditionellen Informationsverarbeitung. Im Zusammenhang mit der Künstlichen Intelligenz werden in diesem Umfeld formal-logische Verfahren entwickelt und angewendet, die Erkenntnis- und Entscheidungsprozesse des Menschen zum Gegenstand haben: Dieser Ansatz wird im folgenden von Cremers/Eder/Hinze im Zusammenhang mit seiner geistesgeschichtlichen Tradition und aktuellen Ausprägungen dargestellt. Dabei wird deutlich, wie die Anforderungen der Künstlichen Intelligenz die mathematisch-logische Forschung und entsprechende technologische Entwicklungen und Anwendungen prägen. Unter dem Aspekt der Symbolverarbeitung kann diese Konzeptquelle als ein wesentliches Element der sogenannten „harten" (symbolistischen) Ausprägungen der KI gelten.

2) Die Biologie als Konzeptquelle

Für die Orientierung von Künstlicher Intelligenz an der Biologie spielen besonders die Motorik (für die Robotik), die Neurophysiologie (Schwerpunkt Hirnforschung) und die Sinnesphysiologie von Menschen und Tieren eine Rolle. Die konnektionistische Ausprägunge der KI (oder Neuroinformatik) basiert wesentlich auf dieser Konzeptquelle. Die Neuroinformatik wird auch als alternatives, konkurrierendes Konzept zu der „harten", symbolorientierten KI verstanden. Dieser Unterschied prägt sich aus im Gegensatz „Neuroinformatik versus KI", wobei hier ein enges, an der Symbolverarbeitung orientiertes Konzept von Künstlicher Intelligenz gemeint ist. Der Beitrag von Eckmiller stellt die Hirnforschung als Konzeptquelle für die Ausprägungen der Künstlichen Intelligenz dar unter Einbeziehung evolutionstheoretischer Aspekte des Menschenbilds.

3) Die technisch-physikalische Realisierung als Konzeptquelle

Die technisch-physikalische Realisierung von Konzepten der Künstlichen Intelligenz im weiten Sinn unterscheidet sich nicht von der technisch-physikalischen Realisierung von Konzepten der traditionellen Informationsverarbeitung. Der Ablauf dieser zwei Formen der Informationsverarbeitung unterscheidet sich auf technisch-physikalischer Ebene auch nicht durch neuronale Architekturen, das Einbeziehen von sogenannter „Fuzzy Logic" oder den Einsatz von Photonenrechner im Gegensatz oder ergänzend zu Elektronenrechnern. Schlachetzki stellt dieses Thema dar. Grundlegend für die technisch-physikalische Realisierung von beiden Formen der Informationsverarbeitung ist ein deterministisches, binäres Modell. Diese Konzeptquelle läßt den Standpunkt plausibel erscheinen, daß Künstliche Intelligenz vom Standpunkt des Entwicklers wesentlich bestimmte Software Engineering Konzepte sind.

4) Die Psychologie als Konzeptquelle

Genau wie die mathematische Logik ist diese Konzeptquelle besonders in den „harten", symbolistischen Ausprägungen der Künstlichen Intelligenz vorzufin-

den. Andrerseits ist eine vielfältige Kritik seitens der Psychologie an Ansätzen der Künstlichen Intelligenz formuliert worden. Verschiedene Konzepte der Psychologie finden sich jedoch zweifelsohne in der Künstlichen Intelligenz wieder. Der Vielfalt der Konzepte von Intelligenz in der Psychologie entspricht eine Vielfalt von Konzepten der Künstlichen Intelligenz. Der Beitrag von Strube thematisiert die Übertragung von Konzepten der Psychologie in die Ausprägungen der Künstlichen Intelligenz unter einem kritischen Blickwinkel. Komplementär zu dieser Betrachtungsweise stellt Wender die Rolle der Computersimulation in der Psychologie dar. Hierbei wird die Leistungsfähigkeit von Computersimulationen von kognitiven Prozessen kritisch bewertet.

5) Integrierte Anwendungsorientierung als Konzeptquelle

Die Orientierung der Künstlichen Intelligenz an der integrierten Anwendung ist im Gegensatz zu den anderen Konzeptquellen nicht eine Orientierung an abstrakten technischen oder wissenschaftlichen Leitvorstellungen, sondern thematisiert den Verwendungszusammenhang der intendierten Anwendung. Diese Ausprägung der KI ist letztlich benutzer- und marktorientiert und kann verschiedene Aspekte der zuvor dargestellten Konzeptquellen integrieren. Der Verwendungszusammenhang kann sich auf Entwickler beziehen, die leistungsfähigere Werkzeuge für die Programmerstellung benutzen wollen, oder auf Anwender, die eine Leistungssteigerung ihrer Anwendung wünschen, die die Orientierung der maschinellen Informationsverarbeitung am Menschen zur Basis hat. Der Beitrag von Müller betrachtet die Einbettung von Konzepten der Künstlichen Intelligenz im Arbeitszusammenhang von Unternehmen und stellt das Konzept der Verteilten Intelligenz als alternative Konzeptquelle und alternative Leitvorstellung zur Künstlichen Intelligenz dar. Haberbeck stellt die Perspektiven der Anwendung und Technologiefolgenabschätzung der Künstlichen Intelligenz dar. In diesem Zusammenhang wird eine integrierte, anwendungsorientierte Leitvorstellung und Konzeptquelle postuliert als Alternative zu abstrakten technischen und wissenschaftlichen Leitvorstellungen und Konzeptquellen.

Die Beiträge in diesem Abschnitt stellen Überlegungen dar, die im Hinblick auf die Technologiefolgenabschätzung im Bereich der KI relevant sind. Es handelt sich jedoch nicht bereits um die Ergebnisse einer umfassenden Technologiefolgenabschätzung. Dies ist die Aufgabe zukünftiger Ausschüsse und Projekte.

Die Rolle der mathematischen Logik in der Künstlichen Intelligenz

Armin B. Cremers, Elmar Eder und Ralf Hinze

Die Logik ist eine der ältesten Wissenschaften, die wir kennen. Die erste schriftliche Überlieferung einer tiefergehenden Behandlung der Logik ist im Organon von Aristoteles zu finden. Bereits dort begegnet uns der für die Logik so zentrale Begriff der Inferenzregel in Form der dort eingeführten Syllogismen. Unter einer Inferenzregel verstehen wir eine Regel, die es erlaubt, aus ein oder mehr wahren Aussagen eine neue wahre Aussage zu erschließen. Syllogismen sind Inferenzregeln, die eine spezielle, von Aristoteles festgelegte, Struktur haben. Als Beispiel sei hier der Syllogismus mit dem Namen Barbara genannt.

Wenn jedes M ein L ist und jedes S ein M ist, dann ist jedes S ein L.

Diese Inferenzregel kann etwa benutzt werden, um aus den wahren Aussagen „Jedes Kind ist ein Mensch" und „Jeder Bub ist ein Kind" die neue Aussage „Jeder Bub ist ein Mensch" zu erschließen.

Mindestens ebenso wichtig für die Entwicklung der Logik wie die Einführung der Inferenzregeln war die Entwicklung einer formalen Sprache der Logik. Es ist das Verdienst von Leibniz, die Vision eines Gebäudes der Logik, bestehend aus einer universellen logischen Sprache (lingua characteristica), einem Inferenzkalkül (calculus rationator) und einer Enzyklopädie des Wissens, formuliert zu haben. Für den recht begrenzten Bereich der Aussagenlogik wurde dies dann von Boole geleistet. Die erste Sprache der Logik, die stark genug war, um darin praktisch die gesamte Mathematik formulieren zu können, war die Begriffsschrift von Frege, die auch heute noch in leicht abgeänderter Form als klassische Prädikatenlogik die am meisten verwendete Logik ist. Bis zur Zeit Freges wurden mathematische Beweise umgangssprachlich formuliert, und dabei wurde an die Intuition des Lesers appelliert. Dies führte gelegentlich zu Widersprüchen dort, wo die menschliche Intuition versagte. Die wesentliche Leistung von Frege bestand darin, daß er eine formale Sprache für die For-

malisierung und zum objektiven Nachvollziehen (Beweisen) von logischen Argumenten schuf. Dieses Nachvollziehen kann rein mechanisch und unabhängig von der Intuition erfolgen. Dadurch konnten solche Widersprüche vermieden werden. Für das Auffinden von Beweisen kann gleichwohl auf die menschliche Intuition nicht verzichtet werden.

Eine der wichtigsten Rollen, die die Sprache der mathematischen Logik in der künstlichen Intelligenz spielt, ist die als Wissensrepräsentationssprache. Die mathematische Logik ist als Mittel zum Ausdrücken von mathematischen Sachverhalten und von logischem Schließen in der Mathematik entwickelt worden. Solches mathematisch präzisierbares Wissen läßt sich in der Logik sehr effizient darstellen. Nehmen wir als Beispiel die obige Aussage „Jedes Kind ist ein Mensch". In der modernen Sprache der Prädikatenlogik wird die Aussage durch die Formel

$$\forall\, x\ (\text{Kind}(x) \rightarrow \text{Mensch}(x))$$

repräsentiert. Wenn wir nun wissen, daß Max, Sabine und Karl Kinder sind, dann können wir daraus logisch schließen, daß sie auch Menschen sind, ohne dieses Wissen gesondert repräsentieren zu müssen. Würde man anstatt dessen eine explizite Darstellung aller Kinder und, gesondert, aller Menschen, etwa in einer relationalen Datenbank, verwenden, so müßte man nicht nur eine Menge unnötiger weil redundanter Information abspeichern, sondern der Zusammenhang zwischen den Begriffen „Kind" und „Mensch" würde verlorengehen. Das Erkennen und Ausnutzen solcher abstrakter Zusammenhänge ist aber gerade eine wesentliche Voraussetzung für intelligentes Handeln. Die mathematische Logik zeichnet sich gegenüber anderen für die Wissensrepräsentation vorgeschlagenen Sprachen gerade durch ihre Ausdrucksstärke und Flexibilität bei der Formulierung abstrakter Zusammenhänge aus. Sie wird daher auf dem Gebiet der Wissensrepräsentation immer mehr Bedeutung gewinnen, je mehr die KI-Systeme in der Lage sein werden, auch mit komplexen abstrakten Zusammenhängen umzugehen.

Ebenso wichtig wie die Darstellung des Wissens in KI-Systemen ist das Erschließen von neuem Wissen aus bereits bekanntem Wissen mit Hilfe von Inferenzregeln. Solche Inferenzmechanismen sind die zentralen Bestandteile

von wissensbasierten Systemen, von automatischen Theorembeweisern und von Systemen zur logischen Programmierung (z.B. Prolog), aber auch von Systemen zur automatischen Programmsynthese, von Systemen zum Umgang mit natürlicher Sprache, und von vielen anderen KI-Systemen. Sie erlauben es erst, die Ausdrucksstärke, die die Logik zur Verfügung stellt, auch auszunutzen. Gerade hier wird aber auch ein Problem heutiger logikbasierter Systeme deutlich. Die Ausdrucksstärke der Logik hat zur Folge, daß die Suchräume außerordentlich groß werden und daher aus Effizienzgründen eine drastische Einschränkung der Suchmöglichkeiten erforderlich ist. Meist wird bereits die Sprache der Logik so eingeschränkt, daß sie leichter handhabbar wird. Beispiele dafür sind die Klauselogik in der Resolution, dem heute am weitesten verbreiteten Kalkül zum automatischen logischen Schließen, sowie die Hornklauselogik in der logischen Programmierung. Die meisten Sprachen zur Wissensrepräsentation sind nichts anderes als eingeschränkte Formen der Prädikatenlogik. Hierzu gehören etwa semantische Netze und Frames. Solche eingeschränkte Sprachen erlauben in ihrem engeren Anwendungsgebiet oft eine einfachere und effizientere Implementierung, als dies bei Verwendung der vollen Sprache der Prädikatenlogik möglich wäre.

Ein wesentliches Merkmal der klassischen Prädikatenlogik ist die Unabhängigkeit der Wahrheit einer Aussage von der Wahrheit anderer Aussagen. Wenn im Rahmen der Lösung einer Aufgabe mehrere Aussagen über das zu behandelnde Gebiet gemacht und nachgewiesen werden müssen, so kann der Nachweis der Wahrheit einer dieser Aussagen unabhängig vom Nachweis der Wahrheit der übrigen Aussagen geschehen. Hierdurch eignet sich die Logik besonders für die parallele Lösung von Aufgaben auf Multirechnersystemen. Als universelle Sprache kann die Logik dabei als Sprache für die Schnittstelle zwischen den einzelnen Rechnern oder Prozessen dienen. Jeder der Rechner liefert Aussagen, die jeder andere dieser Rechner selbst wieder als wahre Aussagen weiterverwenden kann. Die Wahrheit einer Aussage hängt dabei nicht von dem Rechner ab, der sie nachgewiesen hat oder der sie weiterverwendet. Es ist allerdings eine Abhängigkeit zwischen den einzelnen Prozessen insofern vorhanden, als der Beweis einer Aussage für die Lösung des Problems nur dann sinnvoll ist, wenn sich diese Aussage mit den anderen zu beweisenden Aussagen in geeigneter Weise kombinieren läßt, um zu einer Lösung des Problems zu gelangen. Es muß also eine Kompatibilität der einzelnen Aussagen

in Bezug auf ihre Kombinierbarkeit gewährleistet werden, und dazu ist eine Kommunikation zwischen den einzelnen Prozessen nötig.

Die Grenzen der auf klassischer Prädikatenlogik basierten Repräsentation und Verarbeitung von Wissen werden deutlich, wenn man sich klar macht, für welche Anwendungen diese Logik entwickelt wurde. Frege ging es dabei um die formale Darstellung mathematischer Begriffe und Sätze sowie um einen formalen Kalkül, in dem ein Mathematiker Beweise von Sätzen exakt formulieren kann. Das ins Auge gefaßte Anwendungsgebiet war also in erster Linie die Mathematik. Daher eignet sich die klassische Logik zum mathematisch exakten logischen Schließen. Die ersten Logiksysteme und auch die meisten heute gebräuchlichen Logiksysteme haben also zwei Voraussetzungen für ihre Anwendung. Erstens muß das Anwendungsgebiet mathematisch formalisierbar sein. Zweitens muß ein Mathematiker oder sonst ein intelligentes System vorhanden sein, das in der Lage ist, Beweise hinreichend schnell zu finden.

Zunächst zum ersten Punkt. Die klassische Logik als Konzeptquelle reicht nicht aus etwa für das Schließen mit Wahrscheinlichkeitsaussagen oder mit Aussagen über Wissen oder Glauben. Hierzu sind andere Logiken oder Theorien über solches Schließen nötig. Beispiele hierfür sind Fuzzy Sets und autoepistemisches Schließen. Schwierigkeiten, die beim Schließen über Wissen auftreten können, werden etwa in dem folgenden bekannten Paradoxon deutlich.

Einem zum Tode Verurteilten wird mitgeteilt, daß er nächste Woche hingerichtet wird, daß er aber bis zum Zeitpunkt der Hinrichtung den Tag der Hinrichtung nicht wissen wird. Er schließt nun, daß er nicht am Sonntag hingerichtet werden kann, da ja der Sonntag der letzte Tag der Woche ist und er den Hinrichtungstag bereits am Samstag Abend wissen würde. Da er dies weiß, kann er auch nicht am Samstag hingerichtet werden, da er sonst bereits am Freitag Abend den Hinrichtungstag wissen würde. So schließt er schließlich jeden Tag der Woche aus und ist sich sicher, daß er überhaupt nicht hingerichtet wird. Schließlich wird er am Donnerstag zur Hinrichtung geführt, ohne daß er vorher den Tag gewußt hätte.

Ein weiteres Beispiel für die Grenzen klassischer Logik ist das nicht-monotone Schließen. Wissen im täglichen Leben ist selten auf alle möglichen Fälle

anwendbar. Oft hat man Aussagen wie „Alle Vögel können fliegen" zusammen mit Aussagen wie „Pinguine können nicht fliegen", und man muß mit solchen Inkonsistenzen fertig werden. Hierzu werden in der KI Systeme des nichtmonotonen Schließens und Default-Logiken eingesetzt. Intuitionistische Logik wird eingesetzt, wo es darum geht, konstruktive Beweise zu führen. Dies bedeutet, daß man bei einem Existenzbeweis nicht nur die Existenz eines Objektes mit gegebenen Eigenschaften nachweist, sondern daß man dieses Objekt auch explizit angibt. Erweiterungen der klassischen Logik, die häufig benutzt werden, sind die Modallogiken und Temporallogiken. Es ist zwar in der klassischen Logik möglich, etwa die Zeit explizit als Parameter zu verwenden, aber oft ist es zweckmäßiger, dafür eine spezielle Temporallogik zu verwenden.

Aus dem oben Gesagten ist erkennbar, daß klassische Logik gut geeignet ist, um *Situationen* zu beschreiben. Ihre Schwächen werden offenbar, wenn man versucht, *Aktionen* zu formalisieren, wie etwa chemische Reaktionsgleichungen.

$$2H_2 + O_2 \rightarrow H_2O$$

Diese Gleichung kann nicht als logische Implikation gedeutet werden, da als Ergebnis der chemischen Reaktion die Moleküle auf der linken Seite verbraucht werden – klassische Logik beschreibt stabile Wahrheiten, die logische Implikation ist nicht kausal.

Vielversprechend ist in diesem Zusammenhang die lineare Logik, mit deren Hilfe die Verfügbarkeit beschränkter Ressourcen, aber auch die Revision deklarativer Wissensbasen adäquat formalisiert werden kann. Durch Forschungen, die eine Integration der klassischen, intuitionistischen und linearen Logik anstreben, wird diesem Gebiet eine zunehmende Bedeutung zugemessen.

Nun zum zweiten Punkt. Ein automatischer Beweiser geht beim Beweisen eines Satzes so vor, daß er in einem formalen Beweiskalkül nacheinander Schritte ausführt, bis ein Beweis gefunden ist. Ein Mathematiker geht aber ganz anders vor. Er verwendet seine Intuition und erzeugt sich zunächst oft eine grobe Beweisidee, die er dann erst in einem mehr oder weniger formalen Kalkül konkret ausführt. Solange es nicht gelingt, diese Vorgehensweise auch im Computer in befriedigender Weise nachzuvollziehen, wird aufgrund des gewaltigen

Suchraumes die Anwendung des automatischen logischen Schließens höchstwahrscheinlich auf Aufgabenbereiche beschränkt bleiben, die keine große begriffliche Komplexität aufweisen. Bereits in der Aussagenlogik ist die Frage der Gültigkeit einer Formel so komplex, daß sie nach heutigem Wissen jenseits der Behandelbarkeit mit einem Computer ist. In der Informatik wird diese Art von Komplexität als co-NP-Vollständigkeit bezeichnet. Es besteht die starke Vermutung, daß co-NP-vollständige Probleme sich generell und in alle Zukunft nicht mit realistischem Zeitaufwand lösen lassen. Dies bedeutet allerdings nicht, daß nicht auch heute schon für eingeschränkte und trotzdem interessante Klassen von Formeln Beweise effizient mit dem Computer zu führen sind.

Neuroinformatik und Künstliche Intelligenz

R. Eckmiller

Ausgangspunkte von Neuroinformatik und KI

Neuroinformatik und Künstliche Intelligenz sind Teilgebiete der Informatik. Beide Teilgebiete streben danach, typische Informationsverarbeitungs-Leistungen von Menschen und Tieren (Mustererkennung, Navigation, Bewegungssteuerung, Prädiktion, etc.) in technischen Informationsverarbeitungs-Systemen zu erbringen.

Die Software-Ansätze der Künstlichen Intelligenz basieren vor allem auf Erkenntnissen und Modellvorstellungen von Psychologie und Kognitionswissenschaften, also dem Bemühen von Menschen (unter Verwendung ihrer biologischen neuronalen Systeme), menschliche Leistungen und Verhaltensweisen ganzheitlich oder modellhaft ohne genaue Bezugnahme auf die physikalischen und informationsverarbeitenden Eigenschaften seiner Funktions-Module zu beschreiben. Mit anderen Worten wird das Modell, welches sich ein beobachtender Mensch von der ‚intelligenten' Leistung eines anderen Menschen (oder Tieres) macht, direkt in einen als Software darstellbaren Algorithmus umgesetzt.

Im Gegensatz hierzu ist die wichtigste Konzeptquelle der Neuroinformatik die Hirnforschung, also die Gesamtheit von neurobiologischem (Physiologie, Anatomie, Biochemie, Biophysik) Wissen über Struktur und Funktion biologischer neuronaler Systeme und seiner Elemente. Demgemäß kann die Neuroinformatik etwas unscharf definiert werden als: Übertragung von Konzepten der Hirnfunktion auf technische Informationsverarbeitungs-Systeme. Vor diesem Hintergrund kann René Descartes als Vater der Neuroinformatik (nicht aber der Künstlichen Intelligenz !) betrachtet werden, da er (Descartes, 1632) in seinem Buch ‚De Homine' in einem Kapitel mit dem Untertitel: ‚Wie eine Maschine gestaltet sein

müßte, die unserem Körper ähnlich ist' in aller Ausführlichkeit versuchte, auch Hirnfunktionen aus der ‚Maschinen-Perspektive' zu erklären.

Etwa gleichzeitig mit der Entwicklung Software-getriebener Computer während des 2. Weltkrieges nahm auch der Ansatz der Neuroinformatik seinen Aufschwung unter dem Mantel der Biokybernetik, Biophysik und Hirnforschung. Anfang der 60er Jahre gab es einen harten Konkurrenzkampf in den USA zwischen dem ‚Special-Software' Ansatz der Künstlichen Intelligenz und dem ‚Special-Hardware' Ansatz der Neuroinformatik, der (jedenfalls in den USA) dazu führte, daß bis Anfang der 80er Jahre nahezu exklusiv auf Künstliche Intelligenz und ihre Konzeptquellen gesetzt wurde. Erst seit einigen Jahren ist die Neuroinformatik wieder in Schwung gekommen (in den USA als Neural Networks for Computing oder als Connectionism verbreitet). Was waren die Gründe für diese ‚Wiedergeburt' eines Forschungsansatzes und welche Konsequenzen ergeben sich für das zugehörige Menschenbild?

Ursachen der ‚Widergeburt' des Forschungsansatzes der Neuroinformatik

- Nach mehr als 20-jähriger massiver Förderung der Künstlichen Intelligenz in den USA (und mit üblicher Verzögerung in Europa) war sowohl bei den Geldgebern, als auch bei den Wissenschaftlern eine deutliche Ernüchterung bezüglich der technisch vorzeigbaren ‚intelligenten' Funktionen (über die Ebene von Expertensystemen hinaus) eingetreten und führte zu grundsätzlicher Kritik an Konzeptquellen und Ansätzen.

- In der Hirnforschung hatte sich eine Große Menge an neuen Erkenntnissen bezüglich Struktur und Funktion von Nervenzellen, Synapsen und Nervenzell-Verbänden angesammelt, die nun bereit stand,aus der Informatik-Perspektive durchgearbeitet zu werden.

- Es wurden interessante Querverbindungen zwischen Assoziativ-Speichern, dem Verhalten bestimmter Metall-Legierungen (Spin-Gläser) und den Modellen lernfähiger biologischer Nervenzell-Verbände hergestellt.

- Die technische Realisierung von mehr als 1 Million Prozessor -Schaltungen, die in den 60er Jahren unvorstellbar war, rückte jetzt in Reichweite der Mikroelektronik.

- Es wurde darauf hingewiesen, daß z.B. die Fähigkeit einer Fliege mit weniger als 1 Million Nervenzellen eine Hindernis-vermeidende Flugbahn in Echtzeit (z.B. in einem Blumengeschäft) erfolgreich zu erzeugen, weder mit Ansätzen der Künstlichen Intelligenz erklärt, noch mit den schnellsten gegenwärtig verfügbaren Software-getriebenen Supercomputern funktionell kopiert werden kann.

Tatsächlich spielt der biologische Existenz-Beweis in der gegenwärtigen Diskussion über Aussichten und Schwächen der Neuroinformatik eine wichtige Rolle.

Verkürzt lautet das Argument so: ‚'Die vielfältige Existenz biologischer Systeme, deren Leistungen im wesentlichen ihrem Zentralnervensystem (Gehirn) zugeschrieben werden, beweist, daß im Prinzip mit einem großen Prozessor-Netzwerk (neuronales Netz) in Echtzeit diverse intelligente Funktionen bzw. senso-motorische Abbildungsfunktionen erbracht werden können. Die Neuroinformatik befaßt sich mit der sicher (?) mit verfügbaren mathematischen und experimentellen Methoden zu leistenden Aufklärung der funktionellen und konzeptionellen Details''.

Konsequenzen für das Menschenbild

Das Menschenbild der Neuroinformatik ist etwa deckungsgleich mit dem der Hirnforschung. Einerseits nimmt auch ein Arzt ‚'Reparaturen" an einem Automaten vor, so wie dies ein Mechaniker am Automobil tut. Beide verlassen sich vollständig auf ihr naturwissenschaftliches Fachwissen von dem jeweiligen Automaten, ohne daß z.B. der Arzt noch einen Theologen hinzuzöge. Andererseits leben wir Menschen alle in einer Welt, die regional stark verschiedenen Religionen und Philosophien, nämlich den in Büchern überlieferten Modellen zuneigt. Darüberhinaus sind wir regional stark unterschiedlichen Gruppen-Mei-

nungen (Politik) unterworfen und sind ferner typischerweise einem besonderen Chauvinismus der Species ‚Homo sapiens' verpflichtet, die beinhaltet, daß wir das Maß aller Dinge seien. Dies alles sind im Menschenbild der Neuroinformatik Ergebnisse des ständig aktiven Gehirns. Das Gehirn kann hierbei als eine Förderation vieler neuronaler-Netz-Module betrachtet werden.

Die Neuroinformatik kann sich wegen der Tatsache, daß die allermeisten Ergebnisse der Hirnforschung nicht vom Menschen, sondern von Tieren stammen (Affe, Katze, Maus, Frosch, Fliege, Schnecke, usw.) am Dualismus vorbeimogeln. Andererseits hat in der Hirnforschung insgesamt der Dualismus nur eine verschwindend kleine Anhängerschaft.

Eine der wichtigsten Konsequenzen für das Menschenbild aus der Neuroinformatik-Perspektive besteht vielleicht in der Chance, Tabus zu überwinden, eine neue Sicht des Zusammenhanges zwischen Mensch und Technik zu gewinnen und vor allem völlig neue ‚Weltbilder' und Modelle von sich selbst zu entwickeln.

Tabus

Ich erfahre mich als: Lebewesen, Geschöpf, Automat, Maschine. Dieses gemischte Menschenbild (hier Geschöpf – da Maschine) ist die Konsequenz aus der These, daß sämtliche Vorgänge in der belebten und unbelebten Natur den gleichen Gesetzen folgen, von denen ich wegen meines sehr beschränkten mentalen Fassungsvermögens aber eine nur sehr eingeschränkte und möglicherweise durch die speziellen Informationsverarbeitungsabläufe meines Gehirns verzerrte Kenntnis habe.

Technik

Die Schöpfung baut Automaten, die Lebewesen genannt werden. Einige dieser bezüglich Informationsverarbeitungs-Fähigkeiten besonders leistungsfähigen Lebewesen (nämlich Menschen) bauen Automaten, die sie Maschinen nennen.

Diese Maschinen, ja die gesamte Technik ist in der Schöpfung bereits mit angelegt und dient der Vergrößerung des Handlungs-Repertoires der Menschen. Wir müssen also lernen, diese Handlungs-erweiternden Geräte sinnvoll einzusetzen, so wie wir gelernt haben, uns mit einem Küchenmesser nicht zu verletzen.

Neue Weltbilder und Modell-Vorstellungen = Menschenbilder

Die größte ‚Errungenschaft' der Menschheit, nämlich die Sprache, stellt gleichzeitig auch eines der größten Hindernisse zur Entdeckung neuer Blickwinkel bezüglich der Phänomene: Natur, Mensch etc. dar. Wir sind immer wieder (historisch vielfach belegbar) gefangen in scheinbar unangreifbaren Argumentations-Linien und ‚logischen' Begründungen, warum diese und nur diese Betrachtungsweise, Begründung usw. ‚richtig' sein kann. Bis dann ein Umschwung des Zeitgeistes eine völlig neue ‚Party Line', also allgemein akzeptierte Anschauung bewirkt.

Es ist für mich denkbar, daß meine Annahme, ein Automat mit ‚eingebauten' Zielvorgaben im Rahmen eines ‚Universal-Planes' zu sein, nicht beängstigend, sondern vielmehr befreiend wirken kann. Aus der Sicht dieses Menschenbildes eines von der Hirnforschung stark geprägten Neuroinformatikers ergibt sich etwa folgende Lage:

- Kategorische Verneinung eines philosophischen Dualismus.
- Annahme, daß alle Lebewesen Ergebnisse eines dauernd wirkenden Universal-Planes sind.
- Annahme, daß in jedem Lebewesen ständig das ‚Bemühen' abläuft, in Interaktion mit der Umwelt über diverse Sinnesorgane bestimmte Funktions-Parameter zu optimieren.
- Annahme, daß Menschen zum Zweck der Parameter-Optimierung in besonderem Maße ‚Erfahrungen' und ‚Erkenntnisse' anderer Menschen in Optimierungs-Planung einbeziehen und zu diesem Zweck in besonderem Maße Erfahrungs- und Erkenntnis-Austausch mit anderen Menschen durch Zeichen-Kommunikation betreiben.

- Annahme, daß auch Tiere Zeichen-Kommunikation zum Erfahrungs- und Erkenntnis-Austausch betreiben.
- Annahme, daß sowohl die Optimierungs-Ziele (heute: Reichtum, morgen: Freiheit), als auch die Bewertung der Erfahrungen und Erkenntnisse anderer Menschen immer wieder Neu-Orientierungen erfordern oder erzwingen, was auf ‚Individualisten' eine völlig andere Wirkung hat als auf ‚Massen-Menschen'.
- Annahme, daß Menschen im Prinzip die Fähigkeit haben, ihre ja sehr begrenzte Erkenntnis-Fähigkeit (kleine mentale Leselupe) durch ständige Verstellung der Brennweite (Zoom) zwischen sehr präziser Detail-Arbeit und sehr distanzierter extra-terrestrischer Perspektive sinnvoll einzusetzen.
- Annahme, daß gegenwärtig sehr viele als ‚Wahrheiten' angebotene Optimierungsziele (z.B. aus den Bereichen: Religion, Politik, Philosophie, Psychologie) unerträgliche Mängel haben, die sich unter anderem durch Starrheit, Intoleranz und Mangel an Zukunftsvisionen zeigen.
- Annahme, daß es im Prinzip möglich (und Teil des Universal-Planes) ist, die vielen ‚lokalen' Optimierungs-Ziele von Einzel-Menschen und Gesellschaften langfristig miteinander in Einklang zu bringen, was einer Abstimmung (Synchronisation) lokaler Optimierungs-Ziele mit dem Universal-Plan gleichkommt. Diese Annahme impliziert, daß die Menschheit in Harmonie mit der übrigen Natur vor allem nach *einer* Philosophie, *einem* Weltbild, *einer* Religion suchen muß, die dem Universal-Plan entspricht.
- Annahme, daß das Menschenbild der Neuroinformatik wesentlich dabei helfen kann, die vielen widersprüchlichen ‚lokalen' Optimierungs-Ziele als ‚Neben-Gipfel' zu erkennen und (z.B. durch extra-terrestrische, also sehr distanzierte Perspektive) alle Optimierungs-Anstrengungen auf den ‚Haupt-Gipfel' (Harmonie mit dem Universal-Plan) zu richten.

Künstliche Intelligenz und ihre technisch-physikalische Realisierung

A. Schlachetzki

Einleitung

Wenn man heute intelligente Leistungen, wie sie gemeinhin dem Menschen zugesprochen werden, mit einem technischen System erbringen will, dann kann dafür eigentlich nur der elektronische Rechner (Computer) in der ein oder anderen Ausprägung verwendet werden. Wir wollen in diesem Beitrag der Frage nachgehen, welches die grundlegenden Prinzipien sind, nach denen ein Rechner arbeitet. Inwieweit wir damit im Verständnis dessen, was menschliche Intelligenz im Grunde ausmacht, weitergekommen sind, muß offen bleiben. Wir können allerdings sagen, daß wir die eine bestimmte intelligente Leistung technisch nachgebildet haben.

Wenn man sich dem Verständnis des Phänomens der menschlichen Intelligenz nähert, sind zwei Zugangsweisen zu erkennen. Die erste wollen wir die *symbolische* Ebene nennen. Es stehen Vorstellungen im Vordergrund, die man summarisch mit dem Begriff *Person*, mit *Wissensverarbeitung*, *Mustererkennung* oder ähnlichem umschreiben kann. Es geht darum, menschliche Leistungen und Verhaltensweisen ganzheitlich oder modellhaft zu erfassen, ohne daß genau die Funktion des Systems Mensch und seiner Funktionsmodule geklärt zu sein brauchen. Aus dieser Sicht ist die Grundannahme, daß intelligente Leistungen von Rechnern erbracht werden können, die algorithmische Musterverarbeitung betreiben. Sensoren liefern Eingabedaten, die auf Muster abgebildet werden sowie ineinander nach gewissen Regeln umgeformt und verarbeitet werden. Das Resultat wird genutzt um motorischen „Output" zu steuern, also Aktoren zu bewegen.

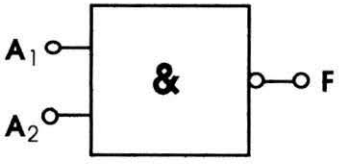
Hier wird der Mensch als System aufgefaßt, das nach erkennbaren Regeln funktioniert und dessen intelligente Leistungen daher auch von geeigneten Automaten erbracht werden können. Aus dieser Sicht wird der Mensch mit einem schwarzen Kasten (black box) gleichgesetzt, dessen Innenleben im einzelnen nicht interessiert, den man aber mit den heutigen technischen Mitteln durch einen elektronischen Rechner repräsentiert.

Der skizzierte Zugang ist gewissermaßen „von oben herab". Seine Beziehung zu der zweiten Zugangsweise, die wir die *funktionale* Ebene nennen wollen, ist nicht klar, auch wenn beide Wege schließlich zum elektronischen Rechner als dem heute dominanten Medium führen. Auf der funktionalen Ebene, mit der wir uns im folgenden befassen wollen, geht es darum, wie man von den kleinsten Komponenten ausgehend schließlich menschliche Hirnfunktionen nachbilden kann. Es handelt sich also um den Weg „von unten nach oben".

Grundzüge der Datenverarbeitung

Aus dieser Sicht ist es zunächst einmal gleichgültig, mit welchen Mitteln wir die erwähnten kleinsten Komponenten oder wie wir den schwarzen Kasten der symbolischen Ebene darstellen. Um praktische Ergebnisse erreichen zu können, tun wir allerdings gut daran, uns an den heutigen technischen Möglichkeiten zu orientieren. Und hier bietet die Halbleitertechnik außerordentlich leistungsfähige Schaltungen, die letztlich auch die bisherige Entwicklung grundlegend geprägt haben. Heutige elektronische Rechner oder die elektronische Datenverarbeitung generell arbeiten fast ausschließlich nach der binären Logik, weil sie sich durch logische Gatter besonders leicht realisieren läßt. Um dies zu veranschaulichen, wählen wir als Beispiel ein logisches Gatter, das weit verbreitet ist. Es handelt sich um die NAND-Funktion, die aus einer Reihe von Transistoren zusammengeschaltet werden kann.Grundsätzlich sind in der binären Logik nur zwei wohl voneinander geschiedene Zustände erlaubt, die wir als logische 0 und logische 1 bezeichnen. Dementsprechend können am Eingang eines NAND–Gatters auch nur diese beiden Zustände erscheinen. Wenn z.B. an beiden Eingängen A1 und A2 jeweils die logische 0 anliegen, dann liefert das NAND–Gatter am Ausgang

die logische 1. Wenn als Variante bei A1 die logische 0 und bei A2 die logische 1



A_1	A_2	F
0	0	1
0	1	1
1	0	1
1	1	0

Fig. 1: Die NAND-Funktion und ihre Funktionstabelle

eingegeben werden, dann nimmt der Ausgang F ebenfalls die logische 1 an. Und so geht es weiter nach Maßgabe der Funktionstabelle des NAND–Gatters.

Die Grundzüge dessen, was wir für das NAND–Gatter erläutert haben, gelten für alle logischen Gatter in der Vielfalt ihrer Spielarten und Erweiterungen, die heute technisch gebräuchlich sind. Immer handelt es sich um eine deterministische Vorgehensweise, die zumindest im Idealfall ohne Fehler abläuft. Wir verstehen darunter, daß für eine gegebene Kombination der Eingabedaten ein wohldefinierter Wert am Ausgang folgt. Da – wie bei jedem technischen System – auch hier Fehler auftreten können, versucht man, sie möglichst zu eliminieren oder zumindest unter einer tolerierbaren Grenze zu halten. In jedem Fall sind Fehler unerwünschte Störungen des Systems, die es – möglicherweise durch Einbau von Redundanz – zu vermeiden gilt. Heute kann man viele Hunderttausende logischer Gatter auf einer einzigen kleinen Halbleiterscheibe unterbringen und diese Scheiben obendrein miteinander kommunizieren lassen. Da man glaubt, mit logischen Gattern das wesentliche der Neuronen, der Elementarbausteine des menschlichen Hirns erfaßt zu haben, hofft man, in wenigen Jahren Maschinen auf den Markt bringen zu können, die die Leistungsfähigkeit des menschlichen Hirns bieten (Electronic World News, 10. Dez. 1990, S. 6). Denn elektronische Neuronen sind in der heutigen Ausführung etwa tausendmal schneller als menschliche Neuronen. Solche Aussagen mögen gewiß überzogen und naiv erscheinen. Sie fußen aber doch in wesentlichen Teilen auf dem, was man heute von der Funktion der Neuronen weiß.

Der Glaube, daß man mit logischen Gattern menschliche Neuronen nachbilden kann, geht davon aus, daß auch diese digital arbeiten, d.h. zeitliche Folgen von Impulsen verwerten, die noch mit gesonderten Gewichten versehen werden. Elektronische Gatter bilden also die wesentlichen Züge der Neuronen nach, nämlich Schwellwertverhalten und Wichtung der Eingänge, auch wenn sie deren Komplexität bei weitem nicht erreichen. Im Jargon der Mikroelektronik ausgedrückt heißt dies, daß sowohl Eingangs- als auch Ausgangsfächer elektronischer Gatter vergleichsweise gering ist.

Was wir für die binäre Logik, also für den heutigen digitalen elektronischen Rechner erläutert haben, gilt prinzipiell auch für den analogen Rechner, der allerdings auf Grund der technischen Entwicklung in den Hintergrund getreten ist. Im Zusammenhang mit der *Fuzzy Logic*, auf die wir gleich eingehen, mag der Analogrechner, wenn auch in abgewandelter Form eine Renaissance erleben, weil er eine Erhöhung der Rechenleistung bietet. Der Trend geht jedoch nicht zum Analogrechner alter Provenienz, sondern zu einem Analogrechner mit digitalisierten Stützstellen. Während beim digitalen Rechner nur zwei Zustände erlaubt sind, kann der analoge Rechner jeden Wert aus einem bestimmten Bereich verarbeiten. Er ordnet einen bestimmten Wert oder Wertesatz am Eingang einen wohldefinierten Wert an seinem Ausgang zu. Jede Abweichung davon ist ein unerwünschter Fehler.

Auch der optische Rechner, dessen Grundelemente heute im Labor erarbeitet werden, liefert prinzipiell nichts Neues über den elektronischen Rechner hinaus. Indem er mit Lichtstrahlen statt mit elektrischen Impulsen arbeitet, verspricht er größere Leistungsfähigkeit gegenüber dem elektronischen Rechner. Das Grundelement beim optischen Rechner, das optische Gatter, verarbeitet zwei Eingangsstrahlen, um gegebenenfalls ein optisches Ausgangssignal zu erhalten. Man bemüht sich, auf optischem Weg ein Analogon in der Art dessen zu bauen, was wir am Beispiel des NAND-Gatters für den elektronischen Rechner erläutert haben. Bisher sind funktionsfähige Einheiten von rund tausend Gattern erreicht worden, also verglichen mit gängigen elektronischen Rechnern ein sehr bescheidenes Niveau. Die Hoffnungen, die an den optischen Rechner geknüpft werden, liegen in seinem Konzept, das den „von Neumann bottle-neck" vermeidet. Er

erlaubt Parallelverarbeitung und Dreidimensionalität, weil sich Lichtstrahlen unbeeinflußt voneinander durchdringen können. Immer aber arbeitet er deterministisch in dem Sinne, wie wir es für die anderen Rechnerspielarten beschrieben haben.

Fuzzy Logic und Neuronale Netze

In den letzten Jahren sind mit der Fuzzy Logic und den neuronalen Netzen zwei Konzepte vorgeschlagen worden, die mehr bieten. Im ersten Fall handelt es sich um eine Annäherung an die unscharfen, „fusseligen" Begriffe der Alltagslogik, im zweiten soll die menschliche Lernfähigkeit nachgebildet werden. In beiden Fällen ist aber wiederum der elektronische Rechner das Vehikel, mit dem die angestrebten Ziele erreicht werden sollen. Es verwundert daher nicht, wenn auch hier im Grunde nach deterministischen Regeln vorgegangen wird.

Um das Vorgehen der Fuzzy Logic anzudeuten, wählen wir als Beispiel die körperliche Größe. Umgangssprachlich quantifizieren wir die diese Eigenschaft mit Begriffen wie „klein", „mittelgroß" oder „groß". Es macht wenig Sinn, eine scharfe Grenze zwischen diesen Begriffen festzulegen, obwohl dies bei technischen Anwendungen oft notwendig ist. Denn wenn eine solche Grenze 1,75 m ist, würden wir kaum einen Menschen mit 1,76 m Körpermaß als groß gegenüber einem anderen mit nur 1,74 m bezeichnen.

Betrachten wir demgegenüber als Beispiel einen Menschen, der 1,80 m groß ist, dann drückt die Fuzzy Logic diesen Sachverhalt durch die Prozentzahlen a und b aus. Wir haben zu Beginn zwei Kurven definiert, nach denen wir große und mittelgroße Menschen klassifizieren (vgl. Fig. 2). Ebenso haben wir Regeln vereinbart, nach denen z.B. Konfektionsware für große Menschen gefertigt werden soll. In diese Regel gehen dann die ermittelten Zahlen a und b ein, bestimmen also zusammen mit anderen Parametern eindeutig den Ausgang der Rechnung. Anders ausgedrückt gehört der betrachtete Mensch zu $a\%$ zu den mittelgroßen und zu $b\%$ zu den großen Menschen. Das Zahlenpaar a und b, das wir für die

Eigenschaft der körperlichen Größe ermittelt haben, wird nach vereinbarten

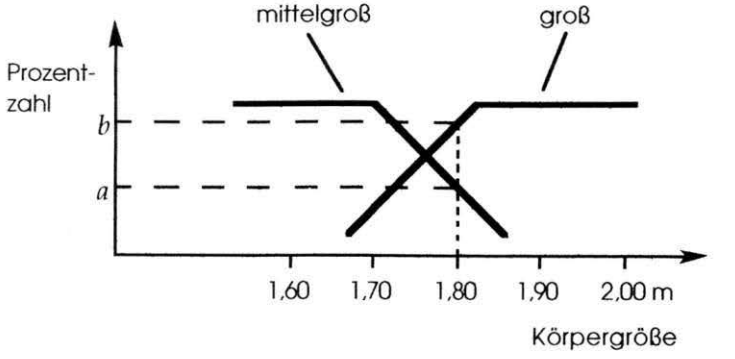


Fig. 2: Zum Vorgehen der Fuzzy Logic

Algorithmen mit entsprechenden Zahlenpaaren für andere Eigenschaften verrechnet mit Ergebnissen, aus denen rückwärts über andere Kurvenfelder konkrete Zahlenwerte für die Schlußfolgerung abgeleitet werden können.

Diese wenigen Andeutungen mögen ausreichen, um zu zeigen, daß die Fuzzy Logic Schattierungen auszudrücken und zu verarbeiten erlaubt, also nicht nur die Eckwerte 0 und 100% (ein Mensch gehört zu *a*% zu den Mittelgroßen und zu *b*% zu den Großen). Im Gegensatz dazu trennt die herkömmliche Logik (Boolesche Algebra) scharf zwischen Ja und Nein (ein Mensch gehört zur Gruppe der Großen: logisch 1; er gehört dann sicher nicht zur Gruppe der Mittelgroßen: logisch 0). In dem Sinne ist die herkömmliche Logik ein Spezialfall der Fuzzy Logic. Diese nutzt wie jene ebenfalls den elektronischen Rechner, braucht i.a. aber mehr Rechenkapazität, was allein schon aus der Notwendigkeit zur Verarbeitung von Kurvenfeldern plausibel wird.

Neuronale Netze bieten eine neue Qualität: sie lernen. Wenn wir ihnen z.B. ein bestimmtes Eingabemuster (e_1, e_2, e_3, ...) anbieten, etwa den Buchstaben A, dann erwarten wir eine wohldefinierte Reaktion, etwa daß das Lämpchen a_1 aufleuchtet. Falls dies nicht geschieht, greift das überwachende System, das auch der Mensch sein kann, ein und ändert das neuronale Netz intern so, daß ein dem

erwarteten Resultat ähnlicheres Ergebnis herauskommt. Es ist eine Frage der Praktikabilität, wann wir uns nach zahlreichen Optimierungsläufen mit dem Resultat zufrieden geben. Wenn die innere Verschaltung des neuronalen Netzes nun soweit optimiert ist, daß das Lämpchen a_3 hell aufleuchtet, wenn wir am Eingang den Buchstaben C als Muster anbieten, a_{15} oder a_{17} aber nur noch schwach glimmen, dann sagen wir, daß das System gelernt hat. Für das Prinzip der Determiniertheit spielt es keine Rolle, ob wir stochastische Ansätze zur Optimierung verwenden, oder nicht.

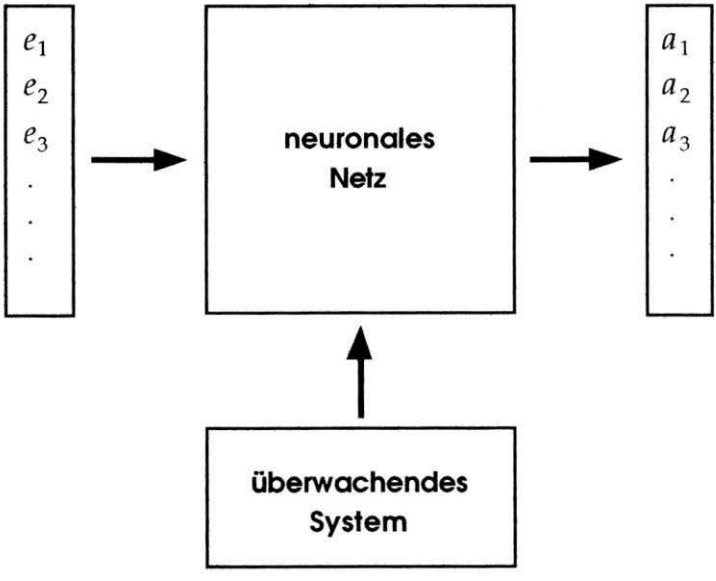


Fig. 3: Lernen mit neuronalen Netzen

Lernen besteht aus der Umorganisation und Umgewichtung der einzelnen Verbindungen innerhalb eines neuronalen Netzes, wobei eine Fülle von Möglichkeiten denkbar ist. Die Verbindungen zwischen den einzelnen Gattern können vorwärtsgerichtet sein; sie können mit Rückkopplung, mit Gewichtung, mit oder ohne Verzögerung und in unterschiedlichem Grad miteinander verbunden sein. Wissen ist verteilt in der Auslegung der Verbindungen repräsentiert. Damit hängt eine erhebliche Toleranz gegenüber Fehlern zusammen: Es macht wenig aus, wenn ein Element ausfällt.

Wir haben bisher vom überwachten Lernen gesprochen, bei dem z.B. der Mensch die Kontrolle ausübt. Wenn wir zum unüberwachten Lernen übergehen, dann ersetzen wir das überwachende System durch einen Satz wohldefinierter Lernregeln. Nach diesen Lernregeln optimiert sich das Gesamtsystem, während wir es sich selbst überlassen. Was im einzelnen im Gesamtsystem während dieses Prozesses vor sich geht, wissen wir nicht, braucht uns auch nicht zu interessieren. Da die technische Realisierung heute mit logischen Gattern der oben geschilderten Art geschieht, ist kaum etwas anderes als ein streng determinierter Ablauf vorstellbar. Die ganz gewiß auch auftretenden Fehler sind unerwünscht und müssen bei einem leistungsfähigeren Nachfolgesystem eliminiert werden.

Fuzzy Logic und neuronale Netze sind nicht direkt miteinander vergleichbar. Der Fuzzy Logic fehlt die Lernfähigkeit, aber Fehlverhalten läßt sich bis auf seine Entstehungsquelle zurückführen. Die Arbeitsweise neuronaler Netze ist auf der Basis der zugrundeliegenden Elemente, etwa von Gattern und einfachen Zusammenschaltungen verstehbar. In größeren Zusammenhängen sind die Vorgänge jedoch nicht mehr nachvollziehbar und daher auch nicht verstanden.

Gegenwärtig sieht es danach aus, als ob sich die beiden erwähnten Techniken verknüpfen lassen, um intelligente Leistungen im hier betrachteten technischen Verständnis nachzubilden.

Zum Verständnis intelligenter Leistungen

Welches Verständnis haben wir nun von intelligenten Leistungen, die von einem denkenden Wesen erbracht werden können? Die diskutierten Beispiele sollten erläutern, welche Realisierungsmöglichkeiten heute gegeben sind, d.h. welche technischen Nachbildungen machbar sind. Diese lehnen sich in vielen Punkten daran an, was man heute von der Funktion der Neuronen weiß. Insofern führen sie mit den heutigen technischen Mitteln genau das aus, was in einem Gehirn bei der Erbringung intelligenter Leistungen vor sich geht.

Alle technischen Realisierungen, die wir bisher besprochen haben, gehen von einem streng deterministischen Ansatz aus. Es liegt ein mechanistisches Weltbild zugrunde, wie es sich in der Physik des 19. Jahrhunderts ausgeprägt darstellt. Daran ändert auch nichts, daß in weitem Umfang Halbleiter und Laser in bewundernswert komplexen Systemen sehr erfolgreich eingesetzt werden und daß sowohl Halbleiter als auch Laser ohne Quantenphysik, die in hohem Maße mit Begriffen wie Unbestimmtheit und Wahrscheinlichkeit arbeitet, nicht zu verstehen sind. Denn diese Bauelemente werden als makroskopische Größen, deren Verhalten mit sehr hoher Wahrscheinlichkeit deterministisch bestimmt ist, betrachtet. Kommt man an die Grenzen des technisch Machbaren, wo das statistische Rauschen zunehmend Einfluß gewinnt, so wird dies als mehr oder weniger tolerierbares Fehlverhalten des Systems in Kauf genommen.

Insbesondere wird der Mensch in seinen intelligenten Leistungen als ein nach Regeln funktionierendes System aufgefaßt, dessen vollständige technische Nachbildung nur eine Frage der Zeit und des Aufwandes ist. Daß mit einer solchen Haltung mit großem Erfolg ausgeprägt intelligente Leistungen vollbracht werden können, wissen wir aus zahllosen Beispielen, die heute Selbstverständlichkeiten sind. Sie reichen von z.T. hochkomplexen Rechnungen im kaufmännischen oder wissenschaftlichen Bereich über die Verwaltung von riesigen Lägern, die Lenkung von Verkehrssystemen wie Eisenbahnen, den Entwurf von Möbeln oder hochkomplexen technischen Geräten, den überragenden elektronischen Schachspieler bis hin zur Steuerung von Weltraumraketen und ausgefeilten Waffensystemen. In allen Fällen werden Rechner eingesetzt, die meist sehr viel besser intelligente Leistungen vollbringen, die bisher dem Menschen vorbehalten waren. Allerdings verstehen wir dabei Intelligenz in einem eingeschränkten, eher bescheidenen Sinn. Wir können uns der Definition anschließen, nach der Intelligenz die Fähigkeit ist, Information im technischen Sinn zu verarbeiten. Man braucht nicht viel Phantasie zu der Annahme, daß elektronische Rechner intelligente Leistungen in noch viel größerem Umfang übernehmen werden, die wir als völlig selbstverständlich akzeptieren werden.

Ob trotz aller Erfolge das mechanistische Bild vom Menschen, das all dem zugrundeliegt, tatsächlich auch zutrifft, ist eine Frage der Überzeugung, die jeder

einzelne für sich entscheiden muß. Ein Beweis in der einen oder anderen Richtung ist bisher jedenfalls noch nicht gelungen.

Ausblick

Von dem Standpunkt der technischen Realisierung her, den wir angenommen haben, sind keine Ansätze zu erkennen, die Phänomene wie Bewußtsein, Gefühl oder Willen auch nur andeutungsweise verständlich machen. Trotz dieser Einschränkung sind mit elektronischen Rechnern erstaunliche intelligente Leistungen erbracht worden. Zweifellos ist gegenwärtig kein Ende der Entwicklung abzusehen, nach der Rechner immer komplexer und leistungsfähiger werden, so daß sie in ihren einzelnen Operationen gar nicht mehr überschaubar sind. Dies gibt jedoch keinen Anlaß, den elektronischen Rechner zu mystifizieren, es sei denn man glaubt, daß durch einen Umschlag von Quantität in Qualität irgendwann einmal etwas prinzipiell Neues entsteht.

Viel naheliegender ist allerdings die Frage nach der Testbarkeit. Je komplexer Rechner werden, umso schwieriger bis praktisch unmöglich wird ihr Test unter allen denkbaren Einsatzbedingungen. Da Rechner, wie jedes andere technische Produkt auch, unter ganz erheblichem Zeitdruck entwickelt werden, kann man leicht Zweifel haben, ob sie hinreichend erprobt sind. Es bleibt ein kaum allgemein lösbares Problem, abzuschätzen, bis zu welchem Risiko wir bereit sein werden, eine Entscheidung dem Rechner zu überlassen.

Unabhängig von solchen Erwägungen allgemeiner Natur wird die Entscheidung darüber, ob die künstliche Intelligenz und ihr wichtigstes Werkzeug, der Rechner, weiterentwickelt werden sollen oder nicht, anderweitig gefällt. Dabei sind Motive der Art dominant, ob der Markt Produkte der künstlichen Intelligenz aufnimmt, ob er zu deren Aufnahme präpariert werden kann oder ob aus Gründen der Macht entsprechendes militärisches Gerät benötigt wird.

Literatur

Altrock, C.V. (1991). Über den Daumen gepeilt. c't Heft 3, 1991, S. 188.

Bieri, P. (1989). In: Pöppel, E. (Hg.): Gehirn und Bewußtsein.Weinheim:VCH Verlagsges.

Kemke, C. (1988). Der neuere Konnektionismus. Informatik Spektrum 11, Heft 143.

Kruse, R. et al. (1991). Modellierung von Vagheit und Unsicherheit – Fuzzy Logik und andere Kalküle. KI Heft 4, 1991, S. 13.

Retti, J. et al. (1986). Artificial Intelligence – Eine Einführung. 2. Aufl, Stuttgart: Teubner.

Zerbst, E.W. (1987). Bionik. Stuttgart: Teubner.

Die Rolle psychologischer Konzepte in der Künstlichen Intelligenz

G. Strube

Beleuchtet wird, in welchem Umfang und auf welchem Wege psychologische und alltagspsychologische Konzepte in die Terminologie der KI Eingang gefunden haben. Unterschieden werden der sprachliche Lapsus und der bloß metaphorische Gebrauch psychologischer Begriffe vom theoretischen Gebrauch, für den die heuristische Funktion und damit die positive Rolle psychologischer Konzepte für die KI aufgezeigt wird.

Einleitung

Daß Informatiker, zumal die Vertreter der Künstlichen Intelligenz (KI), nicht gerade zum peniblen Sprachgebrauch neigen, ist bekannt – weshalb sollten sie sich auch, wenn ihnen Logik und andere formale Sprachen für den präzisen Ausdruck zur Hand sind, mit der Muttersprache quälen, wo doch ohnehin die Zweitsprache Englisch soviel wichtiger ist? Dieser Neigung zur sprachlichen Bequemlichkeit, um nicht zu sagen: Schlamperei, verdanken wir Unwörter wie „Testbett" (statt „Prüfstand", womit das englische *testbed* weit weniger mißverständlich übersetzt wäre), die bemerkenswerte „Daumenregel" (als Synonym zu „Heuristik", wo das Deutsche doch anstelle der angelsächsischen *rule of thumb* die grobe „Faustregel" kennt – oder kannte?), sowie ohne Kenntnis des Englischen nur schwer entschlüsselbare Termini wie die in der Diskussion angeblich Geschichten verstehender Systeme auftauchenden „Haupt- und Nebencharaktere", die als Übersetzung von *main and side characters* (d.h. Haupt und Nebenfiguren) eine besonders gelungene Eindeutschung erfahren haben.Vergessen wir auch nicht, daß beim Prozeß der Unifikation „Patterns" gegeneinander „abgemätscht" („abgematcht": wird auch nicht besser!) werden, daß – dies einem Kurs der letzten KI-Frühjahrsschule entnommen – „ein *deal* ein *goal* ist, das von zwei *Agent*-en ge*shared* wird", und daß schließlich „Künstliche Intelligenz" selbst nicht gerade die gelungenste Wortschöpfung darstellt, wie ihre Vertreter, der

Weiterverwendung des Namens gewiß, bereitwillig konzedieren. Solchermaßen vorsichtig geworden, nähern wir uns der Frage nach der Rolle psychologischer Konzepte, und besonders der von uns allen täglich verwendeten, also alltagspsychologischen Begriffe in der KI.

Vorweg gesagt scheint es mir dabei mehrere Ebenen zu geben, nämlich

- den sprachlichen Lapsus, der, obwohl symptomatisch für die Haltung der Vertreter der KI zum präzisen Ausdruck in einer nicht formalen Sprache, ansonsten ohne tieferen Einfluß auf die Disziplin bleibt und daher eigentlich nur zur Einleitung taugt (siehe oben),
- die alltagsnahe Psychologisierung der technischen Welt (wie vordem die Anthropomorphisierung der unbeherrschten Natur), die freilich für die KI selbst ebenfalls ohne Konsequenzen bleibt, wenngleich sie in der Außendarstellung erhebliche Mißdeutungen provoziert und deshalb kritische Debatten auch innerhalb der KI hervorruft; schließlich
- die Aneignung psychologischer Begriffe, seien sie primär der Alltagssprache entlehnt oder sekundär (als psychologischer Fachbegriff) wieder in diese gelangt, für die eigene Theoriesprache, also für die Definition vonForschungszielen, die Charakterisierung von Vorgehensweisen und endlich für die Einschätzung des Erreichten.

Hier gewinnt die psychologische Terminologie, wie ich gleich nachweisen will,- eigentlich Gewicht für die KI, und das hängt mit deren Zielsetzungen zusammen. Denn selbst Autoren, die nicht (wie Charniak & McDermott in ihrer schätzenswerten Introduction to Artificial Intelligence, 1984, S. 7) als letztgültiges Ziel der KI festhalten, es gelte, „eine Person, oder bescheidener, ein Tier zu bauen", gestehen ein, daß es der KI ja gerade um die technische Realisierung solcher Leistungen geht, die den Menschen artspezifisch auszeichnen – und sei es nur der Konkurrenz halber, um Elaine Richs Bonmot zu zitieren (immerhin der vierte Satz in ihrem weitverbreiteten Lehrbuch von 1983): „KI ist der Versuch, Computern Dinge beizubringen, die Menschen zurZeit noch besser können".

Zunächst aber zur vor-theoretischen Rolle psychologischer Begrifflichkeit, ihrem bloß metaphorischen Gebrauch in der KI. Denn der geht nahtlos über in ihre theoretische Funktion, und oft genug maskiert er sie.

Suggestive Benennungen

Es fällt uns Menschen schwer, Tiere oder auch Maschinen, mit denen wir interagieren, nicht wenigstens gelegentlich wie unsereinen zu behandeln. Wer hätte nie seinen Computer verflucht, oder ihm Handeln zugeschrieben (selbst da, wo der Programmierer nicht wie Joseph Weizenbaum mit seinem berühmt gewordenen Programm ELIZA dies zu provozieren suchte)? Keinesfalls „intelligente" Oberflächenmerkmale können diese Tendenz verstärken, z.B. Sprachausgabe: Vom Computer mit höflicher Frauenstimme um neues Papier für den Drucker ersucht zu werden („your prin-ter is out of pa-per") bringt spätestens nach der dritten gleichförmigen Wiederholung ein „verflixt, ich such doch schon das Papier" hervor. Anthropomorphisierung, in früheren Zeiten auf Naturphänomene beschränkt, ist so universal, daß es auch – und gerade! – den Bereich der Informationstechnik erfaßt. Verstärkt wird diese Tendenz noch durch Einsatzkonzepte, die den „Partner Computer" propagieren.

Noch weiter verstärkt wird die Neigung, von Computer und Programm menschlich zu reden, in der Künstlichen Intelligenz, deren Programme ja gerade menschliches Handeln nachbilden sollen. Die Gefahr liegt dort, wo durch die Benennung fälschlicherweise suggeriert wird, daß dieses Ziel bereits erreicht sei. *Wishful naming* nennt McDermott (1976) diese Tendenz und bringt gleich ein Beispiel: „Erinnern Sie sich noch an GPS? Heute ist das ein farbloser Begriff für ein besonders dummes Programm zur Lösung von Buchstabenrätseln. Aber es bedeutete ursprünglich ‚General Problem Solver', also ‚Programm zur Lösung beliebiger Probleme', und hat damit alle in nutzlose Erregung versetzt und von Wichtigerem abgelenkt."

Selbst harmlosen und als Fachausdrücken klar definierten Begriffen der Künstlichen Intelligenz haften Konnotationen an, die Laien in die Irre führen. Nehmen

wir „Suche" als Beispiel. Wer würde schon annehmen, daß es sich dabei um ein Suchen handelt wie das Tasten eines Blinden im Labyrinth? Wer würde nicht dazu neigen, *best-first search* als mehr anzusehen, als es ist?

Ein besonders typischer Fall ist der des *explanation-based learning*, in deutschsprachiger Literatur meist als „erklärungsbasiertes Lernen" übersetzt. Daß es sich dabei um ein höchst simples Generalisierungsverfahren handelt, das sich in nicht mehr als sechs Prolog-Klauseln programmieren läßt, vermuten selbst diejenigen kaum, die schon die Anfangsgründe der Künstlichen Intelligenz passiert haben. Alles, was das Verfahren macht, ist eine bescheidene Generalisierung, indem einige Konstanten durch variable Parameter ersetzt werden. Dies geschieht anhand einer *domain theory*, worunter man ehrlicherweise nichts weiter erwarten darf als ein paar Aussagen und Regeln. Hier muß ich für meine Freunde aus der Logik natürlich eilends hinzusetzen, daß in der Tat alles, was sich logisch ausdrücken läßt, durch Fakten und Regeln beschrieben werden kann: Trotzdem bleibt *explanation-based learning* weit hinter dem zurück, was man aus menschlicher Interaktion als Lernen durch Erklären kennt, und zwar zumindest deshalb, weil die Darstellung auch kleinster Weltausschnitte in existierenden KI-Systemen von wahrlich erschreckender Dürftigkeit ist, und selbst die naiven Theorien von Laien ungleich komplexer sind.

Die Psychologie selbst, das sei hier gerne zugegeben, kommt diesem Problem auch nicht ganz aus: Ihr stellt es sich als Diskrepanz zwischen alltagsverwurzeltem theoretischen Konstrukt (z.B. „Offenheit") und dessen vergleichsweise magerer Operationalisierung durch Fragebogentests oder ähnliches dar. Und bisweilen wird dem Maximum des Menschenmöglichen ein relatives Minimum des technisch Machbaren gegenübergestellt, wie in Searles berühmter Schrift vom Chinesisch-Zimmer (Searle, 1980), wo der Rechner mit unverstandenen Symbolen hantiert, während menschliches Verstehen, von Searle mit dem Attribut der „Intentionalität" bedacht, offenbar Verständnis einer Tiefe meint, die selbst wir nur unter günstigen Umständen erreichen.

Man sieht also, daß der Gebrauch von Wörtern, die jeder zu kennen glaubt, für Sachverhalte und Verfahrensweisen eines nicht alltäglichen, ja sogar besonders

streng reduzierten technischen Bereiches ziemlichen Schaden stiftenkann. Dies gilt zumal für die öffentliche Diskussion über Künstliche Intelligenz, in der man gerne die alltägliche, konnotativ reichhaltige Bedeutung eines Begriffs unterstellt, obwohl dies eigentlich nicht angeht, so daß als unverschämter und ungerechtfertigter Anspruch der Künstlichen Intelligenz erscheint, was bei näherer Betrachtung nur eine Frage der eingeschränkten fachsprachlichen Bedeutung von Begriffen ist. Wahr ist aber auch, daß vollmundige Werbung und auch Protagonisten der KI in ihren Büchern dem Mißverständnis Vorschub leisten, wenn sie etwa in Aussicht stellen, „Psychologen, Linguisten und Philosophen müssen sich in der Künstlichen Intelligenz auskennen, um verstehen zukönnen, was die Prinzipien der Intelligenz sind" (Winston, 1984, S. 2).

Die Aneignung psychologischer Begrifflichkeit

Ich möchte es umgekehrt wie Winston formulieren: Eine wissenschaftliche Disziplin, eben die „Künstliche Intelligenz", die die „Prinzipien der Intelligenz" aufzudecken trachtet, kann nicht vorübergehen an dem, was Psychologen, Linguisten und Philosophen – letztere seit Jahrhunderten – zu den Fragen von Geist, Erkenntnis, Denken, Intelligenz beigetragen haben. Oder in Anlehnung an Charniak und McDermott (1984): Wenn ein künstlicher Mensch das Leitbild der KI-Forschung sein soll, dann erzwingt dies eine Orientierung (oder wenigstens Seitenblicke) auf reale Menschen und die Psychologie als diejenige Wissenschaft, die für die Untersuchung menschlicher Intelligenz zuständig ist. Wie also sollte Künstliche Intelligenz ohne psychologische Begrifflichkeit auskommen!

Fodor (1975) und Pylyshyn (1984) haben in Arbeiten, die der Grundlegung der Kognitionswissenschaft, also sowohl der theoretischen Psychologie als auch der Künstlichen Intelligenz gelten, mit vielen Argumenten die These belegt, daß diejenige Beschreibungsebene, auf der von mentalen Zuständen, von Zielen, Plänen, und von Handlungen gesprochen wird, die eigentlich kognitive sei. Daß dies zumindest eine unverzichtbare Ebene der begrifflichen Distinktion im Bereich des Kognitiven ist, gestehen mit Smolensky (1988) selbst Vertreter einer Richtung zu, die eher eine subsymbolische oder Mikro-Ebene als die für die Theorie-

bildung maßgebliche ansieht. Die im philosophischen Bereich als die erkenntnistheoretische Position des Funktionalismus bekannte Argumentation Fodors und Pylyshyns weiß ferner für sich geltend zu machen, daß eine Reduktion kognitiver auf (letztlich) physikalische Phänomene nicht bloß unpraktikabel, sondern darüber hinaus prinzipiell unmöglich ist (nämlich wegen der mangelnden Eineindeutigkeit der Zuordnung kognitiver und physikalischer Prozesse). Der langen Rede kurzer Sinn: Die Künstliche Intelligenz kann gar nicht anders denn die begrifflichen Kategorien anlegen, die der menschlichen wie der maschinellen Intelligenz vorgängig sind.

Dasselbe ergibt sich aus einer Perspektive, wie sie in jüngster Zeit prononciert von Searle (1990) vertreten wird. Das von ihm propagierte *connection principle* besagt, daß die eigentlich kognitiven Prozesse – das sind für Searle natürlich diejenigen, denen er „Intentionalität", also so etwas wie Weltbezogenheit, zumißt – im Prinzip dem Bewußtsein zugänglich sind (Searle, 1990, S. 13).

Damit unterscheidet sich Searles Position wesentlich vom Mainstream der Kognitionswissenschaft, die – von der kognitiven Psychologie bis zur Künstlichen Intelligenz – nahezu völlig ohne das Thema „Bewußtsein" auszukommen scheint: gerät doch die Essenz kognitiver Vorgänge durch das Insistieren auf dem bewußt Zugänglichen unter Umständen gar aus dem Bereich des durch objektive Beobachtung und Messung Faßbaren, mithin auch aus dem Bereich der experimentellen Psychologie behavioristischer wie kognitivistischer Ausrichtung. Installiert wird dafür die Begrifflichkeit der Bewußtseinspsychologie, wie sie im späten 19. Jahrhundert dominierte und in der Folgezeit als Alltagspsychologie verbreitet (und dabei breitgetreten) wurde. Dies allerdings muß kein unüberwindlicher Gegensatz sein; vielmehr ist es gerade ein Ziel kognitionswissenschaftlicher Theoriebildung, subjektives Erleben und objektiv beobachtbares Verhalten miteinander in Beziehung zu setzen (s.u.). Im übrigen ist das Bewußtsein auch der systematische Ort, von dem aus Fodor zur Sprache der Gedanken (Language of thought, 1975) gelangt und weiter zu seiner Forderung, die Kategorien mentaler Einstellungen, Zielsetzungen und Handlungen zum begrifflichen Grundinventar der Kognitionswissenschaft zu machen.

Hierbei stellt sich die weitere Frage nach dem Verhältnis solcher „Bewußtseinskategorien", also Handlungen und Ereignissen als Einheiten unserer Welt- und Selbstwahrnehmung, zur Welt „an sich" (soweit diese durch physikalische Messung zugänglich ist) und zu unserem eigenen kognitiven Apparat. Was ist überhaupt Bewußtsein, wieso bedient es sich gerade dieser Kategorien, und wieso nehmen wir unsere mentalen Repräsentationen der Welt „draußen" wahr anstatt „innen", obwohl sie doch in unserem kognitiven System generiert werden? Prinz (1991) beantwortet diese Frage mit dem Hinweis, daß der Zweck ja die Probehandeln ermöglichende, also handlungsadäquate Modellierung situativer Gegebenheiten sei, mithin unser internes Weltmodell diese in Kategorien äußerlich gegebener Objekte und Ereignisse, und unsere Tätigkeit als intentionales im Sinne von auf die Welt gerichtetes Handeln begreift. Die explizite, d.h. dem Bewußtsein zugängliche Repräsentation irgendwelcher interner Verarbeitungsprozesse wäre demgegenüber geradezu dysfunktional.

So wäre – jedenfalls ist dies mein Vorschlag – der Bereich bewußtseinsfähiger Gegebenheiten zu begreifen als die Menge der den allgemeinen kognitiven Prozessen (Speichern und Erinnern, Transformieren und mit anderen in Beziehung setzen) verfügbaren Repräsentationen, in Abgrenzung von bloß lokalverfügbaren, die für Input-Output-Systeme typisch sind, also die „Module" im Sinne Fodors (1983). So betrachtet, ist übrigens Bewußtsein eine

notwendige Funktion allgemeiner und also komplexer kognitiver Systeme. Nur wer – im übrigen ohne guten Grund – voraussetzt, daß ohne Bewußtsein alles genauso funktionieren würde (z.B. Bieri, 1992), sieht sich dem „Rätsel" oder „Problem" des Bewußtseins gegenüber.

Für die Künstliche Intelligenz bedeutet das, ebenfalls nach einer im allgemeinen Sinn handlungsadäquaten Repräsentation der „Umwelt" ihrer Systeme zu streben. (Andere intelligente Systeme, wie z.B. menschliche Benutzer oder andere Maschinen in Multi-Agenten-Systemen, gehören natürlich als besonders auffällige Objekte zu dieser Umwelt dazu.) Die Realität sieht zur Zeit aber so aus, daß unsere Systeme ein überaus reduziertes Handlungsspektrum haben, ja, daß sie typischerweise nur einen einzigen Zweck erfüllen wie etwa den, über eine

Datenbank Auskunft zu geben. Hierfür kann man sich beschränken (und muß es gegenwärtig auch tun), um wenigstens das für diese eine Aufgabe unbedingt Nötige zu repräsentieren. Daraus ergibt sich das Dilemma: Theoretisch ist für die Vertreter der KI und der Kognitionswissenschaft klar, welch hohe Komplexität anzustreben ist, praktisch ist dies gegenwärtig nicht umsetzbar.

Dieses Dilemma führt notwendig dazu, daß der – wie ich hoffe, gezeigt zu haben: notwendig – in den Kategorien der Psychologie zu formulierende Anspruch deutlich überschüssig ist gegenüber der Wirklichkeit vorhandener technischer Realisierungen. Ein Beispiel hierfür aus der Kognitionswissenschaft: Anderson (1983) hat eine anspruchsvolle Modellierung menschlichen Wissens und der darauf operierenden kognitiven Prozesse vorgelegt. Dem von Newell und Simon (1972) für menschliches Problemlösen vorgeschlagenen Modell von Produktionsregeln verpflichtet, ergänzt Anderson diese Kontrollstruktur insbesondere um explizite Zielhierarchien. Aber in der Implementierung bleiben diese Ziele (*goals*) nicht viel mehr als uninterpretierte Symbole. Sie sind einer weiteren Analyse durch das System (z.B. einer semantischen Interpretation) nicht zugänglich und bleiben damit letztlich Etiketten, die lediglich dem menschlichen Leser des Quellcodes Hilfestellung bieten für die Interpretation. Hier ist man also trotz weitergehenden theoretischen Anspruchs realiter immer noch auf der Stufe des *wishful naming*. Auch gegenüber der in der Kognitionswissenschaft verbreiteten Ansicht, wonach Bewußtsein als eine Art Monitorprogramm anzusehen sei (z.B. Johnson-Laird, 1988), gilt meines Erachtens, daß der wesentliche Schritt darin liegt, daß ein aktuelles Ziel nicht nur (als uninterpretiertes Etikett) gemeldet, sondern daß es auch im Kontext der Aufgabenstellung und des Weltmodells interpretiert werden kann.

Es sei nicht verhehlt, daß die Modellierung der Architektur kognitiver Systeme, von Ausnahmen wie der eben erwähnten von Anderson abgesehen, erst in jüngster Zeit zu einem allgemein anerkannten Thema der Künstlichen Intelligenz geworden ist. (Das Bulletin der SIGART-Gruppe der ACM hat jüngst ein ganzes Heft [1991, Nr. 4] dem Thema integrierter kognitiver Architekturen gewidmet.) Den Anstoß für diese Entwicklung hat vor allem die Robotik gegeben, in der inzwischen auch die Einbeziehung dezentraler und nur indirekt kommunizieren-

der Subsysteme, also ein der Zoologie entlehntes Architekturkonzept, diskutiert und in Anfängen realisiert wird (Brooks, 1991). Expertensysteme hingegen, lange Jahre hindurch das kommerzielle Aushängeschild der Künstlichen Intelligenz, sind überwiegend nach dem Schema zentraler Inferenzprozesse über einer deklarativen Wissensbasis organisiert, also in unserem Zusammenhang eher uninteressant.

Am klarsten wird die unausweichliche Psychologisierung der Künstlichen Intelligenz, wenn wir die Zielsetzung betrachten, unter der KI-Systeme (bzw. Systeme, die KI-Komponenten enthalten) konstruiert werden. Je spezifischer der Zweck und das Einsatzkonzept, desto leichter kommt man mit rein technischen, also möglichst einfachen Lösungen aus, die gegenüber kognitiv orientierten Ansätzen sowohl den Vorteil geringeren Konstruktionsaufwandes als auch den höherer Effizienz haben. Doch einer möglichst vielseitigen Verwendbarkeit, auch und gerade unter nur wenig vorhersehbaren Umständen, stehen derartige Lösungen entgegen. Und so werden es letztlich anspruchsvolle Einsatzkonzepte für intelligente Maschinen sein, die eine Hinwendung zu kognitiv adäquaten und damit immer auch aufwendigen Lösungen erzwingen werden. Dies gilt nicht erst für ein (hoffentlich nur abstrakt gemeintes) Ziel wie das, einen künstlichen Menschen bauen zu wollen.

Erst in einem solchen Stadium der Systemkomplexität, wie es für die KI in diesem Jahrhundert nicht mehr zu erwarten ist, wird der mit Begriffen wie „Intention" oder „Handlung" gesetzte Anspruch eingelöst werden können, nämlich daß das so Bezeichnete an Komplexität in Struktur und Dynamik dem zu ähneln beginnt, was wir in der Psychologie unter solchen Namen kennen. Aber der Weg dorthin wird auch interessante Fragen für die Psychologie aufwerfen, und die durch erste Modellansätze angestoßene psychologische Forschung vermag, wie dies auch beim Thema Expertenwissen der Fall gewesen ist, neue Hinweise zur Konstruktion entsprechender Systeme geben. Dies jedenfalls wäre die wünschenswerte Entwicklung. Doch selbst dann gilt, daß vorerst eine deutliche Diskrepanz besteht zwischen der Bedeutung psychologischer Begriffe in der Psychologie und ihren vergleichsweise dürftigen Vorkommensweisen in der Künstlichen Intelligenz. Dieser Diskrepanz sollten wir uns bewußt sein. Den

gegenwärtigen Zustand aber abschaffen – etwa in dem Sinne, daß man der KI verbieten wollte, sich einer psychologischen Begrifflichkeit zu bedienen – wäre meiner Meinung nach kontraproduktiv. Solange Künstliche Intelligenz sich am Vorbild des Menschen orientiert und sich die Konstruktion autonomer und vielseitiger kognitiver Systeme zur Aufgabe stellt, kann sie auf psychologische Begriffe nicht verzichten. Nur sollte sie versuchen, davon präziser zu reden: Aber das brächte uns dann wieder zum Anfang zurück, zum nicht ganz problemlosen und sicher verbesserungsfähigen Verhältnis von Künstlicher Intelligenz und natürlicher Sprache.

Literatur

Anderson, J. R. (1983). The architecture of cognition. Cambridge MA: Harvard Univ. Press.

Bieri, P. (1992, Mai). Was macht das Bewußtsein zu einem Rätsel? Vortrag an der Albert-Ludwigs-Universität Freiburg.

Brooks, R. A. (1991). Intelligence without representation. Artificial Intelligence, 47, 139-159.

Charniak, E.; McDermott, D. (1984). Introduction to Artificial Intelligence. Reading MA: Addison-Wesley.

Fodor, J. A. (1975). The language of thought. New York: Crowell.

Fodor, J. A. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1988). The computer and the mind. Cambridge, MA: Harvard University Press.

McDermott, D. (1976, April). Artificial intelligence meets natural stupidity. SIG-ART Newsletter, 57.

Newell, A.; Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Prinz, W. (1991). Warum wir nicht unser Gehirn sehen. (Paper 11/91). München: Max-Planck-Institut für psychologische Forschung.

Pylyshyn, Z. W. (1984). Computation and cognition. Toward a foundation for cognitive science. Cambridge, MA, London: MIT Press.

Rich, E. (1983). Artificial intelligence. New York: McGraw-Hill.

Searle, J. R. (1980). Minds, brains, and programs. The Behavioral and Brain Sciences, 3, 417-457.

Searle, J. R. (1990). Consciousness, explanatory inversion, and cognitive science. The Behavioral and Brain Sciences, 13, 585-642.

Smolensky, P. (1988). On the proper treatment of connectionism. Behavioral and Brain Sciences, 11, 1-74.

Winston, P. H. (1984). Artificial intelligence (2nd ed.). Reading, MA: Addison-Wesley.

Computersimulation als eine Methode der Psychologie

K. F. Wender

In vielen Bereichen der Psychologie, insbesondere den Teildisziplinen, die menschliches Lernen und Denken untersuchen, ist seit einigen Jahren die Computersimulation eine durchaus populäre Methode. Die Simulation wird hauptsächlich als Werkzeug bei der Theorienbildung betrachtet. Eine empirische Überprüfung solcher Theorien bedeutet dann einen Vergleich von menschlichem Verhalten mit dem Ablauf eines Simulationsprogramms. Der vorliegende Beitrag behandelt die Frage, welche Methoden für einen solchen Vergleich zur Verfügung stehen, welche Schwierigkeiten auftauchen und inwieweit es bis heute gelungen ist, nennenswerte Teile menschlichen Verhaltens zu simulieren. Als Ergebnis wird festgestellt, daß die Simulation einige Prinzipien menschlichen Verhaltens nachbilden kann, daß es aber aus heutiger Sicht als unentscheidbar erscheint, ob der Mensch in irgendeinem Sinne vollständig simuliert werden kann.

Einleitung

Seit einiger Zeit wird immer wieder von Rechnersimulationen als einer Methode gesprochen, die auch in der Psychologie Verwendung findet. Es geht dabei um die Simulation menschlichen Verhaltens, also darum, einen Rechner so zu programmieren, daß er sich so, oder so ähnlich wie ein Mensch verhält. Es ist hiermit nicht unbedingt ein Roboter gemeint, der sich in einer Umgebung bewegt und mit Gegenständen agiert. „Verhalten" meint hier hauptsächlich verbales Verhalten. Das heißt, daß der Rechner Sprache aufnimmt, verarbeitet, in irgendeinem Sinne versteht und seinerseits wieder sprachliche Ausgaben produziert. So hat man zum Beispiel Programme geschrieben, die eine kurze Geschichte, wie etwa eine Fabel, verarbeiten, diese Fabel „verstehen" und Fragen zum Inhalt der Fabel beantworten. Oder man denke an Schachprogramme, denen jeweils eine Stellung eingegeben wird und die mit dem nächsten Zug antworten. Die Frage des Ein- und Ausgabemediums ist dabei eher zweitrangig. Heute werden zumeist Tastatur und Bildschirm verwendet. Wir werden aber wohl nicht mehr allzu lange warten müssen, bis die Ein- und Ausgabe auch akustisch erfolgen kann.

Das Vorhaben, menschliches Verhalten durch Rechnerprogramme nachzubilden, hat ganz offensichtlich etwas mit dem dabei relevanten Menschenbild zu tun. Der vorliegende Beitrag will die Möglichkeiten und Grenzen dieses Vorhabens darstellen. Die bei der Simulation verwendeten Methoden sind im Großen und Ganzen dem Bereich der künstlichen Intelligenz zuzurechnen. Insofern ist es angebracht, Fragen der Simulation menschlichen Verhaltens in einem Buch über „Das Menschenbild in der KI" zu diskutieren.

Die Psychologie hat im Vergleich zu anderen Wissenschaften noch eine vergleichsweise junge Geschichte. Jenachdem, wo genau man die Anfänge sieht, spricht man von etwa einhundert Jahren. Im Mainstream hat sich die Psychologie seither als ein empirisches, naturwissenschaftliches Fach verstanden. In dieser Zeit wurde eine große Menge an empirischen Ergebnissen gesammelt. Es wird jedoch immer wieder beklagt, daß die psychologische Theorienbildung hinter der Vielfalt empirischer Befunde, zumindest im Vergleich mit anderen Naturwissenschaften, zurückbleibt. Dieser Auffassung wird man insbesondere dann zustimmen müssen, wenn man von Theorien einen gewissen Präzisionsgrad verlangt.

Psychologische Theorien sind durchweg verbal formuliert. Diese verbalen Formulierungen verwenden mehr oder minder genau definierte Fachausdrücke sowie im Wesentlichen die Alltagssprache. Durch die Verwendung der Alltagssprache ergibt sich oftmals eine gewisse Unschärfe der Begriffe. Ein notorisch schillernder Begriff ist zum Beispiel das „Unbewußte". Diese begriffliche Unschärfe bedingt, daß die Theorien einen unbefriedigenden Grad an Präzision aufweisen. Präzision in der Theorienbildung wird in anderen Disziplinen durch eine Formalisierung erreicht, die meist mit einer Mathematisierung einhergeht. In dem Maße, wie gesetzmäßige Beziehungen durch mathematische Ausdrücke beschrieben werden, steigt die Prä?zision der Theorien.

Anwendungen mathematischer Methoden gibt es natürlich auch schon lange in der Psychologie. Hierzu zählt insbesondere die Anwendung statistischer Verfahren, die stellenweise sehr weit entwickelt worden sind. Der Einsatz statistischer Verfahren ist jedoch für sich genommen noch keine psychologische Theorie. Statistik ist bei der Auswertung von Experimenten unverzichtbar und sie präzisiert

auch die Planung. Statistische Verfahren können nicht an die Stelle psychologischer Hypothesenbildung oder psychologischer Theorien treten.

Teilgebiete der Psychologie, insbesondere die Mathematische Psychologie, entwickeln auch weitergehende Formalisierungen psychologischer Theorien. Insbesondere die Verwendung von Wahrscheinlichkeitsprozessen ist hier erfolgreich gewesen. Betrachtet man allerdings den Einfluß und die Breite dieser Teildisziplin, so muß man einräumen, daß bisher im Vergleich wenige und nur recht spezielle Bereiche fruchtbar bearbeitet worden sind. Es handelt sich dabei zumeist um Modellierungen von Verhaltensweisen, die unter vergleichsweise eingeschränkten und gut kontrollierten Laborsituationen erhoben wurden. Hierzu gehören etwa Untersuchungen zur visuellen Wahrnehmung oder des Kurzzeitgedächtnisses. Dabei wurden wesentliche und grundlegende Kenntnisse gewonnen. Eine exakte Modellierung war aber nur für recht spezielle Datensätze und nicht für komplexeres Verhalten möglich, wie es etwa schon das Lösen einer eingekleideten Rechenaufgabe darstellt.

Als Verallgemeinerung ergibt sich: Die experimentelle Psychologie hat eine kaum überblickbare Zahl von Beobachtungen und eine Fülle empirischer und auch reproduzierbarer Ergebnisse gefunden. Auch theoretische Konzeptionen auf einem allgemeinen Niveau sind in großer Zahl entwickelt worden. Aber exakt formulierte Gesetze oder Modelle gibt es nur in wenigen Bereichen.

Angesichts dieser Situation sind drei Konsequenzen möglich: Entweder man schließt sich dem Lager derjenigen an, die behaupten, der Mensch und menschliches Verhalten lasse sich sowieso und grundsätzlich nicht formal oder mathematisch beschreiben oder man gibt nicht auf und beschreitet den bisherigen Weg weiter oder man sucht nach neuen Beschreibungsmöglichkeiten. Für diesen dritten Weg ist die Rechnersimulation eine Alternative, die seit einiger Zeit Beachtung gefunden hat.

Rechnersimulationen komplexen Verhaltens

In komplexen Situationen, in denen es eine große Zahl von (nichtlinearen) Einflüssen und wechselseitigen Abhängigkeiten gibt, ist es oftmals nicht möglich, das Verhalten in übersichtlicher, mathematischer Form zu beschreiben. Solche Situationen gibt es z.B. in der Meteorologie oder der Ökonometrie. Hier bedient man sich dann der Rechnersimulation, d.h. man schreibt Programme, die das komplexe Verhalten möglichst gut nachahmen. Ähnliches wird auch in der Psychologie seit einigen Jahren versucht.

Bei diesem Vorhaben erfährt die Psychologie Unterstützung von zwei Nachbardisziplinen: Der Philosophie und der Informatik, speziell der künstlichen Intelligenz. Aus benachbarten Teilen dieser drei Fächer, zusammen mit Teilen der Linguistik und den Neurowissenschaften, entwickelt sich seit einigen Jahren ein neues Fach: Die Kognitionswissenschaft (Cognitive Science). Ziel dieser Unternehmung ist es, informationsverarbeitende Prozesse im weitesten Sinne zu erforschen (Habel, Kanngießer & Strube, 1990).

Wie in anderen Abschnitten des vorliegenden Buches dargestellt, sind in der Künstlichen Intelligenz Methoden der Wissensrepräsentation, des Erlernens, der Verarbeitung und des Abrufs von Wissen erfunden worden. Diese Verfahren haben teilweise Eingang in die Psychologie gefunden und wurden dort als Konstrukte der Theorienbildung eingesetzt. Hierauf wird unten noch eingegangen.

In der Philosophie wurden mit dem Funktionalismus Grundlagen zur Untersuchung kognitiver Prozesse gelegt (vergleiche hierzu das Kapitel des vorliegenden Buches von Kemmerling). Man geht davon aus, daß mentale Zustände nichts anderes sind als physikalische Zustände, die in ihrer Funktion, d.h. funktional beschrieben werden können. Newell (1980) entwickelte den Begriff eines Physical Symbol System und argumentiert, daß es bezüglich der Funktionen nicht auf die physikalische Realisierung ankomme. Kognitive Vorgänge werden als ein informationsverarbeitendes System angesehen, das aus Repräsentationen besteht, auf welchen Prozesse arbeiten, die funktional beschrieben werden kön-

nen. Zur Begründung dieser Argumentation sei z. B. auf das Buch von Pylyshyn (1984) hingewiesen.

Verbunden mit der Auffassung von mentalen Vorgängen als informationsverarbeitende Prozesse ist das Vorhaben, diese Vorgänge durch Programme zu beschreiben. Dies ist der Ansatzpunkt von Rechnersimulationen in der Psychologie. Insbesondere seit symbolverarbeitende Programmiersprachen wie LISP oder Prolog entwickelt wurden, hat man vielfältige Möglichkeiten gesehen, mentale Prozesse durch Programme zu beschreiben.

Gemeinsame Konzepte in Künstlicher Intelligenz und Psychologie

Die Psychologie hat in ihrer Geschichte wiederholt Anregungen oder Analogien aus anderen Wissenschaften als Konstrukte der Theorienbildung übernommen. Beispielsweise die in der Linguistik von Chomsky entwickelte Phrasenstrukturgrammatik wurde als Modell auch des menschlichen Sprachverstehens angesehen. Oder die Konzeption des menschlichen Kurzzeitgedächtnisses als ein Schieberegister ist eine Anleihe aus der Elektrotechnik.

Auch mit der Künstlichen Intelligenz hat die Psychologie viele Konzepte gemeinsam. Dabei ist in vielen Fällen nicht ganz klar, in welchem Fach, Künstliche Intelligenz oder Psychologie, ein bestimmtes Konzept zuerst verwendet wurde. Zu den gemeinsamen Konzepten gehören unter anderem: Semantische Netze, symbolische Suchräume, heuristische Suche, Schemata, Skripte und Frames (Rahmen), Produktionsysteme, Übergangsnetzwerke und Marker Passing.

Semantische Netzwerke in verschiedenen Varianten sind eine oftmals verwendete Methode der Wissensrepräsentation. Ein semantisches Netz wurde zuerst von Quillian (1968) sowie Collins und Quillian (1969) als Modell für das menschliche Langzeitgedächtnis konzipiert. Die von ihnen postulierte hierarchische Struktur war mit einer Intersektionssuche kombiniert, die später zum Marker Passing erweitert wurde. Vorhersagen des ursprünglichen Modells stellten sich in Experimenten als zu restriktiv heraus, so daß erweiterte Formen entwik-

kelt wurden. Seither sind semantische Netze in der einen oder anderen Variante Bestandteil vieler psychologischer Theorien, die menschliche Gedächtnisleistungen beschreiben.

Newell und Simon (1972) gehörten zu den ersten, die menschliches Problemlösen durch ein Rechnerprogramm simulierten. (Simon erhielt später den Nobelpreis.) Newell und Simon analysierten Problemlösesituationen als heuristische Suche in symbolischen Suchräumen (Problem Space). Problemlösen wurde als das Finden eines Weges von einem gegebenen Ausgangszustand in einen erwünschten Endzustand definiert. Der von Newell und Simon entwickelte General Problem Solver konnte vermittels der Means-Ends-Analysis Probleme in verschiedenen Denksportaufgaben und bei einfachen logischen Schlüssen Lösungen finden. Der General Problem Solver war explizit als ein Programm zur Simulation menschlichen Denkens konzipiert. Allerdings hat sich die Hoffnung, daß er sich über die zunächst behandelten Probleme hinaus verallgemeinern lassen würde, nicht bewahrheitet.

Produktionssysteme (Post, 1943) wurden in der Psychologie ebenfalls von Newell (1973b) zur Simulation von Suchprozessen im Kurzzeitgedächtnis eingesetzt. Vergleichsweise sehr umfangreiche Systeme, die ganz oder teilweise aus Produktionssystemen bestehen, entwickelten Anderson (1983) und Newell mit Mitarbeitern Laird, Newell, Rosenbloom, (1987). Newell's SOAR ist als System für menschliches Lernen entworfen und Andersons ACT* soll Lernen, Denken und Sprache umfassen. Produktionssysteme sind Teile vieler weiterer psychologischer Simulationsprogramme (Möbus, 1988; Opwis, 1988).

Das Konzept „Frame" wurde wohl zuerst von Minsky vorgeschlagen. Die speziellere Variante „Script" entwickelten Schank und Abelson (1977) im Rahmen ihrer Arbeiten zur Verarbeitung natürlicher Sprache. Auch dieses Konzept wurde in die Psychologie übernommen. Es regte eine Reihe experimenteller Arbeiten an. Frames sind eine Form der Wissensrepräsentation, die sich seither in vielen Zusammenhängen bewährt hat.

Der theoretische Status von Simulationen

Man muß zunächst zwischen zwei unterschiedlichen Ansätzen unterscheiden. Soll ein Programm eine bestimmte Aufgabe erfüllen wobei zunächst beliebig ist, auf welchem Wege dies geschieht oder soll das Programm die Aufgabe möglichst genauso wie Menschen, d. h. auf dem gleichen Wege, mit den gleichen Teilprozessen lösen? In die erste Kategorie gehören z. B. Schachprogramme. Sie „spielen" inzwischen hervorragend Schach aber sie wurden meist nicht so entwickelt, daß sie einen menschlichen Spieler dabei nachahmen. Hier würde man im engeren Sinne nicht von Simulation sprechen. Anders ist es beispielsweise beim General Problem Solver. Dieses Programm wurde von vornherein so konzipiert, daß es auch den Vorgang und nicht nur das Ergebnis menschlichen Problemlösens nachbildet. Programme dieser zweiten Kategorie sind der eigentliche Gegenstand des vorliegenden Beitrags.

Welchen Status haben nun Simulationen in der Psychologie? Kann ein Simulationsprogramm an die Stelle einer Theorie treten? Kann man behaupten: Das Simulationsprogramm ist meine Theorie?

Diese letzte Auffassung läßt sich nicht verteidigen. Sollte jemand ein Programm – ohne dahinterstehende Theorie – geschrieben haben, das tatsächlich einen signifikanten Ausschnitt menschlichen Verhaltens simuliert, so wird das Programm letztlich genauso komplex sein, wie das Verhalten. Hatte man schon keine Theorie zur Erklärung des Verhaltens, dann braucht man jetzt (sozusagen noch zusätzlich) eine Theorie zur Erklärung des Programms.

Somit ist klar, daß Simulationen kein Ersatz für eine Theorie sein können. Von Anhängern der Simulation wird dann auch lediglich behauptet, daß Simulationen die Theorie explizieren. Diese Explikation hat zwei Vorteile. Sie kann erstens mögliche Widersprüche in der Theorie aufdecken (und gewissermaßen einen Existenzbeweis liefern) und zweitens genauere Hypothesen zur empirischen Prüfung ermöglichen.

Simulation als Präzisierung der Theorie

Wenn man behauptet, daß eine Simulation eine Theorie expliziert, dann muß man letztlich beweisen, daß ein Programm tatsächlich die Konzepte einer Theorie und die Folgerungen daraus realisiert. Man müßte also eine möglichst präzise Darstellung der Theorie haben und dann einen möglichst formalen Beweis, daß das Programm auch diese Theorie darstellt.

Solange dies, wie zur Zeit eigentlich bei allen Anwendungen in der Psyvchologie, nicht der Fall ist, bleibt die Möglichkeit der rationalen Rekonstruktion. Damit ist gemeint, daß die Theorie unabhängig, ein zweites Mal in ein Programm übertragen wird und daß ein Erfolg darin besteht, wenn beide Programme identische Ergebnisse liefern. Diese Art der Prüfung ist recht aufwendig und bisher selten durchgeführt worden.

Empirische Prüfungen von Simulationen

Wie steht es nun um die experimentelle oder empirische Prüfung einer Simulation? Das Vorgehen besteht hier allgemein darin, daß Programm gewisse Ausgaben produzieren zu lassen und mit menschlichem Verhalten zu vergleichen. Die Angaben sollen dabei Zwischenergebnissen von Teilprozessen entsprechen. Zwei Aspekte können dabei beachtet werden: Die jeweilige Art der Ausgabe, was also Programm bzw. Mensch gerade tut, und die Zeitabstände zwischen einzelnen Ausgaben.

Hier ergeben sich nun eine Reihe von Fragen, von denen einige erwähnt werden sollen. Das erste Problem ist der Auflösungsgrad. Wie klein sollen die Einzelschritte, Teilhandlungen sein, für die ein Vergleich zwischen Programmausgabe und empirischer Beobachtung durchgeführt wird. Evidenterweise können sie nicht zu global sein, denn dann würde auch die mögliche Übereinstimmung zu global sein und es könnte viele Teilprozesse geben, in denen das Programm (oder die dahinterstehende Theorie) nicht stimmt. Man kann aber auch nicht zu sehr in die Einzelheiten gehen, denn dann wird man erstens nie fertig und zweitens ist

klar, daß ja zumindest die physikalische Realisierung von mentalen Prozessen und Programm unterschiedlich ist und daß sie – dem Funktionalismus folgend – keine Relevanz besitzt. An welcher Stelle hier die Grenze zu ziehen sei, ist unklar. Das Ergebnis einer empirischen Überprüfung einer Simulation hängt aber jedenfalls von dem gewählten Auflösungsgrad ab.

Welche Verhaltensweisen bzw. welche Programmausgaben soll man nun überhaupt zum Vergleich heranziehen? Die Antwort beruht wesentlich darauf, welche Aspekte menschlichen Verhaltens im untersuchten Bereich der registrierenden Beobachtung zugänglich sind. Insbesondere, wenn sogenannte höhere kognitive Prozesse beteiligt sind, ist das keine triviale Frage. Die Psychologie hat einige Methoden entwickelt, hier die Datenbasis zu erweitern. Hierzu gehört die Methode des lauten Denkens und die Aufzeichnung von Blickbewegungen. Es würde zu weit führen, hier auf Einzelheiten dieser Methoden einzugehen. Man muß aber darauf hinweisen, daß bei beiden Methoden im Detail Interpretationsschwierigkeiten bestehen.

Eine weitere Möglichkeit zur Überprüfung besteht darin, Personen Aufgaben am Rechner bearbeiten zu lassen und alle Tastendrucke zu registrieren und als Vergleichsbasis zu nehmen. Aber auch hier wird sofort deutlich, daß Personen natürlich in den Pausen zwischen den Tastendrucken mentale Aktivitäten durchführen können, die somit unbeobachtet bleiben.

Als Ergebnis läßt sich festhalten, daß es schwierig ist zu entscheiden, welche Verhaltensweisen von Personen mit welchen Programmausgaben verglichen werden sollen. Im praktischen Fall sieht man sich gezwungen, gewissermaßen definitorisch Festlegungen vorzunehmen.

Resümee

Zum impliziten Menschenbild der kognitiven Psychologie oder auch der Kognitionswissenschaft gehört die Auffassung, daß sich kognitive Prozesse (zumindest Teile davon) durch Rechnerprogramme simulieren lassen. Es gibt Publikationen,

die diese Möglichkeit für einzelne Bereiche demonstrieren. Es muß aber betont werden, daß eine Simulation nur dann sinnvoll sein kann, wenn sie auf einer psychologischen Theorie basiert. Die Simulation sieht eine Theorie voraus und kann sie gegebenenfalls präzisieren. Eine Simulation kann zeigen, daß eine gegebene Theorie in der Lage, d. h. hinreichend ist, ein bestimmtes Verhalten zu produzieren.

Ähnlich wie in der Mathematischen Psychologie gilt jedoch auf hier, daß sich die Simulationen bislang eher auf begrenzte Bereiche beziehen. Um einen „hinkenden" Vergleich zu bemühen: Man stelle sich vor, daß die Funktion und die Funktionsweise eines menschlichen Beines nachgebildet werden soll. Die Psychologie ist jetzt vielleicht etwa so weit, daß sie eine „Holzbein" (eventuell mit einem Gelenk) entwickelt hat. Der Patient steht, aber bis zu einer annähernd vollständigen Simulation, einschließlich der physiologischen Prozesse, ist es noch ein weiter Weg. Bei der Entwicklung und Überprüfung von Simulationsprogrammen für mentale Prozesse müssen zwei Probleme überwunden werden. Erstens müssen die kognitiven Prozesse näher erforscht werden und zweitens ist die Frage der empirischen Prüfung neu zu fassen.

Aus dem Gesagten soll man nicht schließen, daß durch Methoden der KI nicht intelligentes Verhalten produziert werden kann. Schachprogramme auf Großmeisterniveau sind ein deutliches Beispiel. Nur solche Programme simulieren nicht notwendig menschliche kognitive Prozesse.

Die technischen Neuerungen im Hard- und Softwarebereich vollziehen sich sehr schnell und eröffnen immer neue Möglichkeiten. Programme, die in irgendeinem Sinne intelligentes Verhalten produzieren, werden immer raffinierter. Nur, ob diese Intelligenz – empirisch geprüft – in einem umfassenden Sinne menschlichen mentalen Prozessen entspricht, werden wir sobald nicht erfahren.

Literatur

Anderson, J. R. (1983). The architecture of cognition. Cambridge, Mass.: Harvard University Press.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, Mass.: M. I. T. Press.

Collins, A. M.; Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Habel, C.; Kanngießer, S.; Strube, G. (1990). Editorial. Kognitionswissenschaft, 1:1, S. 1-3.

Laird, J. E.; Newell, A.; Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. Artificial Intelligence, 33, 1-64.

Minsky, M. L. (1985). The society of mind. London: Heinemann.

Möbus, C. (1988). Zur Modellierung kognitiver Prozesse mit daten- bzw. zielorientierten Regelsystemen. In: Mandl, H.; Spada, H. (Hrsg.): Wissenspsychologie. München, Weinheim: Psychologie-Verlags-Union, S. 423-465.

Newell, A. (1973b). Production systems: Models of control structures. In: W. Chase (ed.): Visual information processing. New York: Academic Press.

Newell, A. (1980). Physical symbol systems. Cognitive Science 4:2, 135-183.

Newell, A.; Simon, H. A. (1972). Human problem solving. Englewood Cliffs, N. J.: Prentice-Hall.

Opwis, K. (1988). Produktionssysteme. In: Mandl, H.; Spada, H. (Hrsg.): Wissenspsychologie. München, Weinheim: Psychologie-Verlags-Union, S. 74-98.

Post, E. L. (1943). Formal reductions of the general combinatorical decision problem. American Journal of Mathematics, 65, 197-268.

Pylyshyn, Z. W. (1984). Computation and cognition: Toward a foundation for Cognitive Science. Cambridge, Mass.: Bradford Books, M.I.T. Press.

Quillian, M. R. (1969). The teachable language comprehender: A simulation program and theory of language. Communication of the ACM, 12, 459-476.

Schank R. C.; Abelson, R. (1977). Scripts, plans, goals and understanding. Hillsdale, N. J.: Lawrence Erlbaum Associates Inc.

KI und Menschenbild im Unternehmen*

R. A. Müller

1. Einleitung

Gestalt und Funktion technischer Systeme sind in hohem Maße von Leitbildern der Designer und Entwickler bestimmt. Gegenstand wissenschaftlicher Untersuchungen ist hier z.B. die Frage, ob und wie Leitbilder als Steuerungsinstrumente technischer Innovationen eingesetzt werden können (Dierkes u.a 1992). Leitbilder wirken sich auch unabhängig davon aus, ob sie den Akteuren bewußt sind oder nicht. Auch am Beispiel von Anwendungen der Künstlichen Intelligenz (KI) im Unternehmen zeigt sich, wie Erfolg oder Mißerfolg technischer Systeme im Unternehmen entscheidend von Leitbildern abhängen kann.

Der vorliegende Beitrag konfrontiert das technische Leitbild der klassischen KI (Kap. 2), das gleichermaßen ein Menschenbild ist, mit aktuellen und in weiten Bereichen akzeptierten Unternehmensleitbildern (Kap. 3). Es wird die These aufgestellt, daß der geringe Erfolg bisheriger KI-Anstrengungen damit zusammenhängt, daß das KI-Leitbild mit dem Unternehmensleitbild unverträglich ist. Seit Mitte der achtziger Jahre entstanden daher Konzepte, die vom Leitbild der klassischen KI abwichen und die heute innerhalb der Informationstechnik unter der Bezeichnung „Computer als Medium" (Kap. 4) zusammengefaßt werden können. Der hier beschriebene Ansatz („verteilte Intelligenz" (VI)" und „aktive Medien", Kap. 5.1) läßt sich dieser Strömung zuordnen. Auf der Grundlage dieses VI-Ansatzes wurden inzwischen einige Systeme und Prototypen erfolgreich implementiert und im Feldversuch getestet (Kap. 5.2).

*Wichtige Ideen zum Konzept „Verteilte Intelligenz" verdanke ich meinem Kollegen Alexander Mankowsky.

In den entsprechenden, stark anwendungsorientierten Forschungsprojekten wurden wertvolle Erfahrungen gesammelt, wie Kenntnisse aus der Industriesoziologie und der Arbeitspsychologie nicht nur über den „Umweg" der Leitbildentwicklung sondern auch unmittelbar in die Prozesse des System-Design und -Engineering einfließen. Zur Zeit wird daran gearbeitet, diese („soziotechnische") Vorgehensweise weiter zu verfeinern, zu verallgemeinern und zu validieren, mit dem Ziel, den Effizienz- und Akzeptanzgewinn, der bei den bisher erstellten Systemen erkennbar ist, einer möglichst großen Systemklasse zugute kommen zu lassen.

2. Leitbild und Menschenbild der KI

Das Eigentümliche und gegenüber vielen anderen Technologien neuartige der klassischen KI besteht darin, daß ihr Leitbild mit ihrem Menschenbild identisch ist. Prägnant ist die von Minsky vertretene Position der „harten" KI: „Der Mensch ist eine Fleischmaschine" (differenziert setzen sich mit Positionen der KI Capurro, Seetzen, Kemmerling, Krämer, Becker und Lischka auseinander, 1992; alle in diesem Band; vgl. auch Rammert 1991). In etwas abgeschwächter Form nahm dieses Leit-/Menschenbild als Expertensystemkonzept Anfang der 70er Jahre eine kommerzielle Gestalt an. Noch in den 80er Jahren warb eine Firma – durchaus ernst gemeint – mit dem Slogan „Der Experte kann gehen – Das Expertensystem bleibt" und suggerierte damit ein Potential zur Senkung von Personalkosten sowie eine Technik zur Konservierung von Spezialkenntnissen. Nach einer Kette nicht enden wollender Mißerfolge (s. Coy und Bonsiepen, 1989) hat dieses Konzept in seiner Urform heute nur noch sehr wenige Anhänger; bei vielen ist die Idee gänzlich diskreditiert. Die Bezeichnung Expertensystem wurde häufig ersetzt durch „wissensbasiertes System", was eher auf die speziellen Programmiersprachen und -Stile der KI verweist als auf das ursprüngliche Ziel anspruchsvolle geistige Arbeit (Problemlösen) an den Rechner zu delegieren. Görz sieht bezüglich des ursprünglichen Leitbildes der KI „eine Verschiebung vom Ziel des ‚autonomen' KI-Systems zum Mensch-Maschine-Tandem" (Görz 1991). Coy und Bonsiepen stellen sogar die Frage, ob es sich bei der Konzeption einer autonomen Maschine „überhaupt um ein lohnendes For-

schungsziel handelt", und ob dieses anthropomorphe Leitbild „nicht einfach eine falsche Zielvorgabe suggeriert" (Coy und Bonsiepen 1989, S. 20).

Entgegen den Erwartungen, die noch in den siebziger Jahren geäußert wurden, ist der Markt reiner KI-Produkte und Applikationen marginal geblieben (mit eher sinkender Tendenz). Offenbar ist „autonome Intelligenz" keine Produkt- oder Verfahrenseigenschaft, nach der ein nennenswerter Markt bisher verlangt hat. Expertensysteme können genau das nicht ersetzen, was ein Unternehmen von einem menschlichen Experten (unter anderem) verlangt: Kreativität, Handlungserfolge auch in untypischen Situationen und außerfachliche (z.B. soziale) Kompetenz. Der selbstironische Slogan „if it works it's not AI" ist ernst zu nehmen: Zumindest dem Verfasser ist unter der ohnehin spärlichen Anzahl erfolgreicher, d.h. im produktiven Einsatz befindlicher „Expertensysteme" kein einziges in Reinform bekannt. Immer treten die Funktionen des Problemlösens (wie etwa Diagnose und Konfiguration) gegenüber anderen Funktionen, die z.B. kooperativen Anforderungen des Arbeitsprozesses genügen müssen, in den Hintergrund. KI-Techniken sind typische Vertreter eingebetteter Funktionalitäten. Ihr Anwendungserfolg hängt weniger von der Intelligenz des eingebetteten KI-Systems ab als vom Design und Zusammenwirken des Gesamtsystems. Das Leitbild der KI ordnet sich in der Praxis erfolgreicher Projekte (ganz im Einklang mit konventionellem Software-Engineering) immer dem Leitbild des Gesamtdesigns unter und ist beim fertigen System folglich nicht mehr erkennbar; auch ein erfahrener KI-Programmierer kann als Anwender eines fertigen Softwaresystems nicht beurteilen, ob konventionelle oder KI-Technik bei der Programmierung verwendet wurde.

Zusammenfassend läßt sich sagen: die zweifellos zukunftsweisenden Konzepte und Programmiertechniken der KI sind zwar Ergebnis von Forschungsbemühungen den Menschen nachzubauen (insofern wäre die o.g. Kritik von Coy und Bonsiepen zurückzuweisen); eine dem gleichen Leitbild folgende Anwendungsstrategie des Ersetzens menschlicher Denkleistung durch maschinelle im Unternehmen hat bisher und wohl auch in absehbarer Zukunft keine Aussicht auf Erfolg. Das Menschenbild der klassischen KI spielt in Unternehmen nicht nur keine sonderlich produktive Rolle, es hat sich durch seine eingeengte Sichtweise

des maschinellen Problemlösens für den Innovationstransfer sogar als hinderlich erwiesen (wie zahlreiche gerade wegen dieses Leitbildes gescheiterten Projekte beweisen; vgl. Coy und Bonsiepen 1989; Rammert 1992) und wurde in den Fällen erfolgreicher Projekte von Leitbildern, die aus dem Anwendungsumfeld oder dem Markt stammen, ersetzt. Es kann (zumindest unter marktwirtschaftlichen Bedingungen) keine in Forschungslabors entstandene Technologie an Unternehmen vorbei Verbreitung finden. Innovationen, deren Leitbilder nicht mit Gegebenheiten des Marktes (bei Produkten) oder der Unternehmenskultur (bei Verfahren) in Einklang stehen, haben keine Chance.

Der bei Technologietransfer immer wieder anzutreffende und offenbar nicht auszurottende Grundfehler, eine Methode (wie die KI) isoliert „in der Praxis anzuwenden", ist nicht neu. Er wurde in der Ära des Operations Research begangen und tritt auch bei Software-Entwicklungsprojekten als Konflikt zwischen Anwender (mit Sachkompetenz) und Kerninformatiker (mit Methodenkompetenz) in Erscheinung.

Dieser Fehler läßt sich im allgemeinen durch ein Projektmanagement vermeiden, das gewährleistet, daß von Anfang an die Sicht und der Nutzen (die Leitbilder) der Praktiker im Vordergrund stehen, und die KI- (allgemein: Methoden-) Kompetenz in der dienenden Rolle einer von vielen Methoden in den Hintergrund tritt. Das von der KI-Gemeinde seit Jahrzehnten beklagte „Integrationsproblem", das es mit softwaretechnischen Mitteln zu lösen gälte (aber für das bis heute trotz größter Anstrengungen keine allgemeingültige Lösung in Sicht ist), ist in Wahrheit primär ein Artefakt, der garnicht in Erscheinung träte, wenn der oben beschriebene Grundfehler vermieden würde.

Damit ist aber die Frage noch nicht beantwortet, wie Leitbilder beschaffen sein können, die zu meßbarem Nutzen bei größerer Anwendungsbreite der KI-Technologie im Unternehmen führen. Auf diese Frage versucht der vorliegende Beitrag eine Antwort zu geben. Der Rahmen des möglichen KI-Einsatzes ist hierbei auf Prozesse der Leistungserstellung innerhalb von Unternehmen oder ähnlichen Organisationen (einschließlich Planung, Kontrolle und Verkauf) festgelegt. KI

als Bestandteil von Produkten (Investitions- oder Konsumgüter, Spiele) bleibt außer Betracht.

3. Unternehmensorientierte Leitbilder

Gemäß dem Menschenbild der klassischen („harten") KI, steht die Sicht des Unternehmens als Ansammlung menschlicher Problemlöser, die zunehmend von maschinellen verdrängt werden, im Vordergrund. Diese Sicht bietet kaum Platz für zeitgemäße und differenzierte Einbeziehung unternehmensorientierter Leit-/ Menschenbilder, die dem Stand moderner Betriebswirtschaftslehre bzw. Managementpraxis (z.B. Staehle 1987, Ulrich und Probst 1990) entsprechen. Diese Unverträglichkeit des klassischen KI-Leitbildes mit unternehmensorientierten Leitbildern wird immer deutlicher als selbstgemachtes Problem der KI erkannt. Zu den ersten, die einen Ausweg vorgeschlagen haben, gehören Winograd und Flores (1986).

Im Zentrum unternehmensorientierter Leitbilder steht heute die Forderung nach einem noch nie dagewesenen Grad von Anpassungsfähigkeit großer Organisationen an ständig zunehmende Komplexität und Dynamik ihres Umfeldes (Märkte, Technologien, Umwelt, Ressourcen). Überkommene hierarchische Organisationsstrukturen, strenge Einhaltung von Ressortbefugnissen, taylorisierte Formen der Arbeitsteilung, Abschottungen zur Sicherung eigener Informationsvorsprünge erweisen sich immer mehr als Hemmnisse zu höherer Effizienz. Das heutige Menschenbild, das einem lebendigen anpassungsfähigen Unternehmen gerecht wird, verlangt engagierte, kreative, kontaktfreudige, kooperative und hochflexible Mitarbeiter mit lebenslanger Bereitschaft zum Lernen. Auch das Unternehmen selbst ist ein lebendiger Organismus, eine eigenständige, hochentwickelte Lebensform (vgl. den Begriff des „Erwerbskörpers" bei Hass 1970). Die Haupteigenschaft (Existenzbedingung) dieser Lebensform ist die Fähigkeit zur Selbstorganisation (Probst 1987). Sie kann durch vielfältige Arten von Deformierung (Sklerosen, Wucherungen) beeinträchtigt, gelähmt oder zerstört werden. Das heutige Leitbild des kreativen unternehmerischen Mitarbeiters ist dem Leitbild des neunzehnten Jahrhunderts, das sich eher am funktionierenden Rädchen

einer Maschine anlehnte, nahezu diametral entgegengesetzt. Dies spiegelt sich auch in der Rezeption von Theorien der Selbstorganisations- und Chaosforschung im Rahmen sozial- und organisationswissenschaftlicher Fragestellungen wieder (Haken 1988, 1991; Balck und Kreibich 1991).

Eine im obigen Sinne ideale Organisation wäre nach Toffler in der Lage, ihre inneren Strukturen jeder neuen äußeren Situation sofort („ad hoc") anzupassen. Eine solche „Adhokratie" ist der Gegenpol zur starren Bürokratie und Hierarchie, die der künftigen Dynamik der Umfeldentwicklungen nicht mehr gewachsen sein werden (Malone und Rockart 1991, S.128).

Diese Idealform der Adhokratie läßt sich unter heutigen Voraussetzungen nicht ohne weiteres realisieren. Sie wäre auch nicht wirtschaftlich. Denn eine solche Organisationsstruktur würde einen sehr hohen Aufwand für Kommunikation und Koordination erfordern. Es ist leicht vorstellbar, wie der Aufwand zur ständigen Optimierung von Arbeitsformen und Organisationsstrukturen die operativ zu erbringende Leistung übersteigen und schließlich ersetzen könnte.

Auch in heutigen Unternehmen sind die Koordinations- und Kommunikationskosten hoch (mit steigender Tendenz). Sie sind um so höher, je differenzierter (spezialisierter) Arbeitsteilung und Wechselwirkungen zwischen einzelnen Abteilungen bzw. Aufgaben sind. Es gilt also nicht erst im Rahmen einer entwikkelten „Adhokratie", daß ein hohes Potential zur Steigerung der Wettbewebsfähigkeit einer Organisation durch Erhöhung der Qualität der Kommunikation bei gleichzeitiger Senkung des Aufwands für Koordination zu erschließen ist. Dies ist der Kern aller Bemühungen, die Leistungserstellung im Unternehmen durch neue Formen der Arbeitsteilung (z.B. Gruppenarbeit) zu verbessern. Es geht um die verbesserte Koordination menschlicher Handlungen, um eine „Adhokratie" als situationsgerechte Zusammenführung menschlicher Kompetenzen und know-how's.

Genau dies durch neue Formen der Computerunterstützung zu ermöglichen ist die Zielsetzung der im folgenden Kapitel beschriebenen Medienperspektive der Informationstechnik. Unter dieser Perspektive sind diejenigen Konzepte, Werk-

zeuge und Produkte einzuordnen, mit denen diese Verbesserung angestrebt werden. Die Basisfunktion der elektronischen Post beispielsweise spielt im Rahmen dieser Perspektive neben Datenbanken heute schon eine große und ständig wachsende Rolle. „Electronic mail, which has little to do with distributed computing but everything to do with distributed people, is a fundamental enhancement to options for human communication... It is easy and tempting to think of networks as hooking computers together. They are better thought of as hooking people together, with a computer mediating the connection in an effective way. Whether the application is electronic mail or access to remote information, the motivation for communication is human need, not internals of computer system design." (Clark 1991, S. 31f). Die Besonderheit für Entwickler derartiger Systeme besteht darin, daß sie nicht nur ein Computersystem, sondern ein soziales System entwickeln und demzufolge ihr Hauptaugenmerk auf das Verständnis der Aufgabe und nicht auf die Technologie legen müssen (Turoff 1991, S. 110f). In diesem Aufgabenverständnis tritt der Unterschied zwischen dem Menschenbild der klassischen KI und dem oben skizzierten sehr deutlich zu Tage.

4. Die Medienperspektive der Informationstechnik

In dem Maße wie Produktivitätsreserven durch Automatisierung der Produktionssysteme ausgeschöpft sind, gewinnt die Unterstützung und Automatisierung von Koordinationsfunktionen (Administration und Management im weitesten Sinne) immer mehr an Bedeutung. Der MIT-Report zur Lage der Automobilindustrie (Womack u.a. 1991) verdeutlicht, daß große Chancen in neuen Koordinationsformen der Produktion liegen (z.B. durch Gruppenarbeit). Nach Abschluß der ersten Phase der Automatisierung kontrollieren Menschen nur noch die Automaten, die die Maschinen in der Produktion bedienen; in der nächsten Phase geht es darum, die vorhandenen, meist zentralisierten Formen direkter Administration (Planung, Verteilung von Ressourcen, auch von Informationsressourcen, Verkauf) durch neue computergestützte Formen und Techniken der Koordination zu ersetzen. Nun erweist sich gerade hier die klassische im Unternehmen installierte EDV, wenn sie zu einseitig auf streng sequentielle Arbeitsabläufe ausgerichtet ist, als großer Hemmschuh einer Reorganisation zu mehr Flexibilität. Vorherr-

schend ist hier die vertikale Orientierung der Systeme, indem sie auf den Einzelarbeitsplatz bezogene Aufgaben unterstützen.

Gruppenunterstützung und Teamarbeit erfordern dagegen eine horizontale Orientierung der Unterstützung Einzelarbeitsplatz übergreifender Funktionen (Kommunikation, Koordination). Derartige Unterstützungssysteme zielen auf eine Erhöhung der Kooperationsfähigkeit arbeitsteiliger Teams, auf eine Senkung der Reibungsverluste und folglich der Kosten.

Unter der Medienperspektive lassen sich diejenigen Softwarekonzepte und Systemarchitekturen zusammenfassen, die Hindernisse der Zusammenarbeit über Grenzen von Raum, Zeit und Fachdisziplin hinweg überwinden helfen. Die wachsende Bedeutung dieser Sichtweise spiegelt sich auch in Aktivitäten der Fachgruppe „Computer als Medien" der Gesellschaft für Informatik wieder. Als „Computer-Medien" werden häufig sehr spezielle Konzepte bezeichnet (z.B. Lernhilfen, Multi-Media, Virtual Reality, Hypertext, Electronic Mail,...). Demgegenüber werden hier unter Medien alle auf eine Organisation als soziales System bezogenen Kommunikationsunterstützungs-Funktionen verstanden. Dazu gehören Telefon, E-Mail und Datenbanken als Beispiele passiver elektronischer Medien, aber auch Medien ganz anderer Art wie Kataloge, Stücklisten, Löhne und Preise (zur Regulierung von Austauschprozessen) oder Corporate Identity als Beispiel eines (subtileren) Mediums der Verhaltenssteuerung. Medien überbrücken Kommunikations-Grenzen: Die Einführung von Electronic Mail trägt zur Überwindung von Raumgrenzen bei (analog: Datenbanken bei Zeitgrenzen, Übersetzer bei Sprachgrenzen, Virtual Reality und Tutorielle Systeme bei Kompetenzgrenzen,...). Es genügt aber nicht, die hierfür nötigen softwaretechnischen Plattformen (Datenbanken und Dokumentenaustausch) zu installieren: der weit größere Aufwand besteht darin, diese Systeme um weitere (heute noch fehlende) Komponenten zu ergänzen, die die Effizienz von Arbeitsprozessen, vor allem durch neue Formen der Arbeitsteilung und Integration elektronischer mit nichtelektronischen Medien, erhöhen. Fragen der Entwicklung dieser fehlenden Bestandteile der medialen Infrastruktur sind Forschungsgegenstand sowohl innerhalb der Informationstechnik als auch der Techniksoziologie (Rammert 1989), und gehören heute noch nicht zum Standardrepertoire des Softwareengi-

neering. Jedoch läßt sich bereits heute feststellen, daß bei der Realisierung dieser Medienbestandteile die Techniken der KI eine wichtige Rolle spielen werden. Dies soll im nächsten Kapitel illustriert werden.

Computer sind Universalmaschinen, die sich hinsichtlich ihrer Nutzung und Funktionen unterschiedlich definieren und ausgestalten lassen: z.B. als Rechner, Sortierer, Problemlöser, Mustererkenner, Optimierer, Schreibmaschine, Konstruktionshilfe. Erst in den achtziger Jahren hat sich gegenüber diesen klassischen Auffassungen lokaler, d.h. auf einen einzelnen Arbeitsplatz bezogenen, Funktionen eine Klasse globaler, arbeitsplatzübergreifender Funktionen auch als Leitbild (Computer als Medium) herausgebildet, von dem Konzepte und technische Realisierungen bereits in zahlreichen Spielarten vorliegen, wie z.B. Computer Supported Cooperative Work (CSCW), Organizational Computing (gleichnamige Zeitschrift; Applegate u.a. 1991), Coordination Science (Malone u.a. 1991), Verteilte Intelligenz (Müller 1988, 1991a) und Groupware. Nelius (1992) gibt eine gute Übersicht zum Stand dieser Entwicklungen.

Die heute kommerziell verfügbaren Groupware-Produkte bieten zwar die softwaretechnische Infrastruktur/Plattform zur Vernetzung der einzelnen Arbeitsplätze (Mail, Datenbanken) sowie zur Speicherung und Manipulation von Texten oder Graphiken (z.B. LOTUS NOTES oder COORDINATOR), d.h. einfache, passive Funktionen zur Nachrichtenübermittlung, -speicherung und -wiedergabe. Sie bieten jedoch keine Möglichkeit, z.B. komplexere Modelle (im Sinne virtueller hochdimensionaler Abbildungen der Realität) auf unterschiedlichen Abstraktionsstufen arbeitsteilig zu manipulieren (etwa im Sinne von Concurrent Engineering).

Aus dem Blickwinkel der Medienperspektive sind die bisher realisierten Systeme (z.B. raumüberwindende Medien zur Telearbeit) und Software-Werkzeuge nicht als Endpunkt sondern als Beginn einer neuen Ära medienorientierter Informationstechnik anzusehen. Die Zukunft wird den aktiven Medien gehören, welche z.B. die sachlogischen Bedeutungen von Nachrichten erkennen können, die Systemverträglichkeit von Planungen, Visionen oder Szenarien auf unterschiedlichen Betrachtungsebenen prüfen, Vorschläge verbessern, Nachrichten

aus vorhandenen Bruchstücken synthetisch und sachlich korrekt generieren sowie vor ungewollten Nebeneffekten künftiger Handlungen warnen. Elektronische Märkte werden die Beschaffung und Verteilung von Gütern, Dienstleistungen und Information revolutionieren (Schmid u.a. 1991, Malone und Rockart 1991). Bei Finanzmärkten sind bereits heute bedeutende Transaktionen vollständig auf Computer-Medien verlagert.

Diese Medien müssen keineswegs kompliziert oder gar „intelligent" sein. Wie man aus den Erfahrungen der innerbetrieblichen Nutzung von Expertensystemen weiß, ist Maschinenintelligenz eine Eigenschaften, die den praktischen Einsatz auch behindern kann. Demgegenüber orientiert sich die Gestaltung aktiver Medien am Ideal pflegeleichter, überschaubarer und robuster Informationstechnik, die in ein komplexes, vielschichtiges und fehlertolerantes Netzwerk sozialer Interaktion eingebettet ist. Ihre Effizienz wird allein an der Effizienz der Arbeitsprozesse gemessen und nicht an der rein informationstechnischen Leistung Teilaspekte (z. B. Problemlösen) von Arbeitsprozessen nachzubilden.

5. Verteilte Intelligenz

5.1 Konzeption

Verteilte Intelligenz ist nicht präziser definierbar als der (lokale) Intelligenzbegriff. Es genügt aber für das Verständnis zunächst die folgende grobe Vorstellung: Während sich lokale Intelligenz als Eigenschaft eines individuellen Trägersubjekts verstehen läßt, ist verteilte Intelligenz immer Eigenschaft einer Vielzahl (mindestens zwei) interagierender Individuen.

Es erscheint zweckmäßig, zwischen diesen beiden Sichtweisen von Intelligenz zu unterscheiden; denn einerseits gibt es Organisationen mit drastisch unterschiedlicher Leistungssfähigkeit trotz gleicher Leistungsfähigkeit der Organisationsmitglieder; andererseits können kollektive Fähigkeiten fehlende individuelle Fähigkeiten bis zu einem gewissen Grade kompensieren (Beispiel: Ameisenstaat). Die strukturellen Kopplungen, die das Verhalten als Gruppe oder Organi-

sation determinieren, sind weder von ihrer Entstehung noch ihrer Funktion her als Eigenschaft der Individuen erklärbar. Erfolg und Überlebensfähigkeit einer Organisation sind nicht allein von der Intelligenz der Individuen bestimmt, sondern außerdem von Art und Ausmaß der strukturellen Kopplungen zwischen den Individuen, den internen Kommunikations- und Koordinationsstrukturen.

Auf diesem Intelligenzbegriff baut Verteilte Intelligenz (VI) als ein auf der Ebene von CSCW angesiedeltes informationstechnisches Leitbild auf, das in der Forschung der Daimler-Benz AG entwickelt wurde (Müller 1988, 1991a). Es stimmt in hohem Maße mit den Zielsetzungen von CSCW, Coordination Science oder Organizational Computing (s.o.) überein.

Der VI-Ansatz unterscheidet sich von letzteren vor allem dadurch, daß er die im Kapitel 4 aufgeführten unterschiedlichen Arten von Medien trotz ihrer Heterogenität (elektronische und nichtelektronische Bestandteile) als Einheit betrachtet. Während bei Groupware die Überwindung räumlicher und zeitlicher Beschränkungen von Arbeitsprozessen im Vordergrund steht, konzentriert sich VI auf die Überwindung von Grenzen der „Fachdisziplin". Hiermit ist die wechselseitige Überführung von Modellen, Sichtweisen und Fachsprachen unterschiedlicher Spezialisten bei verketteten Arbeitsprozessen gemeint, z.B. mit Hilfe von Systemen, die eine fertigungsorientierte Produktsicht in eine vertriebs- und kundenorientierte übersetzen.

Ein weiterer wichtiger Unterschied besteht in der Auffassung von „Gruppe". Im CSCW-Rahmen wird hierunter meist eine Art Projektgruppe, ein gezielt installierter Zusammenschluß von Personen zur Bearbeitung einer Aufgabe, verstanden. Beim VI-Konzept ist die Gruppe (allgemeiner) durch die arbeitsteiligen Prozessen der innerbetrieblichen Leistungserstellung definiert. Gruppe kann dort auch eine Konstellation von Personen sein, die sich ihres Gruppencharakters weder bewußt sein noch je räumlich zusammenkommen müssen (z.B. Mitarbeiter aus Produktion, Vertrieb und Kunden). Kommunikation kann völlig anonym stattfinden (z.B. über Datenbanken).

Der Ausdruck „Verteilte Intelligenz" wurde bewußt in Anlehnung an KI gewählt, weil mit der Formulierung des VI-Ansatzes auch der Versuch einer Überwindung des klassischen KI-Leitbildes einherging, ein Versuch die Wirksamkeit von KI-Techniken durch „Interferenz" (Dierkes u.a. 1992) mit einem gültigen Unternehmensleitbild zu erhöhen. Beim VI-Konzept handelt es sich um eine Perspektive, die in der Informatik (und anderen Wissenschaften) allgemein als Wende von einem rein (informations)technischen zu einem mehr soziotechnischen Systemverständnis zu sehen ist (vgl. Coy 1989 und sein Nachwort in der deutschen Ausgabe von Winograd und Flores; Welter 1988; Floyd 1989; Klotz 1991; außerhalb der Informatik: s. z.B. Trist 1981, Probst 1987).

Dem VI-Konzept liegt der Kommunikationsbegriff von Maturana und Varela als handlungskoordinierende Kopplung sozialer Systemkomponenten zugrunde. „Unter Kommunikation verstehen wir dabei das gegenseitige Auslösen von koordinierten Verhaltensweisen unter den Mitgliedern einer sozialen Einheit" (Maturana und Varela 1987, S. 210).

In Technik und Alltag wird Kommunikation meist mit dem Transport von Nachrichten und Dokumenten gleichgesetzt; in Wirklichkeit handelt es sich bei diesem Transport um einen oberflächlichen Teilaspekt von Kommunikation, in Maturanas Theorie sogar um eine „grundfalsche" Auffassung: „Das Phänomen der Kommunikation hängt nicht von dem ab, was übermittelt wird, sondern von dem, was im Empfänger geschieht. Und dies hat wenig zu tun mit ‚übertragener Information'" (Maturana und Varela 1987, S.212).

Neben diesem Kommunikationsbegriff von Maturana und Varela und dem oben (Kap. 3) bereits beschriebenen Leitbild des Unternehmens als selbstorganisierendes System stützt sich das VI-Konzept auf die Systemtheorien von Bunge (1977, 1979) und Luhmann (1988). Besonders das Unternehmensleitbild wirkt sich auf die praktische Realisierung aus. Das auf VI basierte Design führt zu einer größeren Anpassungsfähigkeit der Unterstützungssysteme an kommunikative Erfordernisse der Arbeitsprozesse (s.u. das Pkw-Kunden-Beratungssystem PBES) als z.B. das auf Winograd und Flores (1986) zurückgehende COORDINATOR-

System, bei dem sich Arbeitsprozesse an (rigide) Vorgaben des Systems anpassen müssen (Nelius 1992, S. 38f).

Das herkömmliche Softwareengineering bringt als Ergebnis eines i.a. sehr aufwendigen Analyse-, Design- und Produktionsprozesses relativ komplexe Software mit reichhaltigen Funktionen und im Vorhinein schwer abzuschätzender Benutzerakzeptanz hervor. Unser Ansatz zielt darauf ab, den Analyse- und Designprozeß so zu verändern, daß die Herstellung (im Sinne verteilter Intelligenz) wirkungsvoller Software mit hoher Akzeptanz besser beherrscht wird.

5.2 Praktische Erfahrungen

Das VI-Konzept wurde bisher in zwei parallelen Forschungsprojekten („Beratungs- und Kommunikationssysteme" und „Elektronische Informationsmärkte") durch Entwicklung entsprechender Werkzeuge und Prototypen für spezielle Arbeitsprozesse im Unternehmen umgesetzt. Arbeitsprozesse wie z.B. Verkaufen, Beraten, Planen sind Kommunikationsprozesse in einem selbstorganisierenden System. Diese Kommunikationsprozesse sind eben wegen ihrer Eigenschaft der Selbstorganisation nicht unmittelbar von außen erfolgsorientiert steuerbar. Gestaltbar sind jedoch die Medien dieser Kommunikation.

Im Projekt „Beratungs- und Kommunikationssysteme" geht es um die Unterstützung des Gesprächs zwischen Verkäufer und Kunde. Die Situation ist dadurch gekennzeichnet, daß der Gesprächsverlauf durch die Komplexität des Produkts (Variantenvielfalt und Sonderausstattungen eines Fahrzeugs) bei Verwendung herkömmlicher Medien (Kataloge, Preislisten, telefonische Rückfragen) behindert und daß Detailfragen und Wünschen des Kunden (z.B. bestimmte Qualitäts- und Leistungsanforderungen, Liefertermine, Baubarkeit, Preise oder Finanzierungsart) nicht schnell und verläßlich genug entsprochen werden kann. Es wurde ein aktives Medium entwickelt (auf Lap Top PC), bei dem der Arbeitsprozeß des Verkaufens als Gruppenprozeß, an dem Verkäufer und Kunde beteiligt sind, aufgefaßt wird. Zwei Prototypen PBES und OBES sind im praktischen Einsatz. Die Leitidee der Verteilten Intelligenz kommt im Gesamtdesign sowie zahlreichen

Details zum Ausdruck: Verkäufer und Kunde erleben das System in erster Linie als elektronisches Handbuch das über aktive Hintergrund-Funktionen (Verträglichkeitsprüfung, Berücksichtigung hersteller- und kundenseitiger Strategien) verfügt. Die beim klassischen KI-Ansatz dominierende Funktion des Konfigurierens ist zwar vorhanden, tritt jedoch als eine von mehreren Funktionen nicht explizit in Erscheinung. Sie wirkt sich lediglich implizit als Qualitäts- und Effizienzgewinn des Arbeitsprozesses im Rahmen der kundengesteuerten Auftragserstellung aus.

Aktives Medium heißt im Fall PBES, nicht nur die direkten sondern auch die nicht anwesenden (virtuellen) Kommunikanten im Verkaufsgespräch zu beteiligen, ohne die unmittelbare Kommunikation zwischen Verkäufer und Kunde zu stören. Dies geschieht dadurch, daß Bedingungen, die aus Produktion, Konstruktion oder Finanzierung stammen, in einer Wissensbasis abgelegt sind und automatisch berücksichtigt werden. Akzeptanz und Nutzen von PBES ist in der Funktion begründet, diese Informationen zur richtigen Zeit (prozessuale Sicht) in der richtigen Form in den Arbeitsprozess einfließen zu lassen. Der Programmablauf ist nicht zielgerichtet auf einen Bauauftrag, sondern zyklisch auf den Prozess der Willensbildung hin angelegt. Die Wahlfreiheit für den Kunden (Metapher der „Palette") in jeder Gesprächsphase ist hier das herausragende Systemmerkmal.

Das zweite Projekt „Elektronische Informationsmärkte" bezieht sich auf räumlich und sachlich verteilte Planungs-Prozesse im Unternehmen. Planen ist ein Kommunikationsprozeß, der in mehrerlei Hinsicht verteilt stattfindet:

- Vielzahl handelnder Individuen, Gruppen, Teilsysteme,
- Aufgabenverteilung zwischen Mensch und Computer,
- zeitlich (ungleichzeitig) und räumlich (an mehreren Orten),
- unterschiedliche Spezialisierungen innerhalb der Organisation (Qualifikation, Kompetenz, Know-how).

Planung in diesem Sinne kann u.a. die folgenden Funktionen umfassen:

- Prognose von Marktpotentialen (lokal, global, in Abhängigkeit von der Wirtschafts- und Sozialstruktur),
- Konsistenzprüfung und -sicherung unternehmerischer Zielsetzungen,
- Controlling einer vernetzten Händlerorganisation,
- Optimierung der betrieblichen Leistungserstellung.

„Elektronischer Markt" ist ein neuartiges softwaretechnisches Funktionskonzept, das als aktives Medium in unterschiedlichen Varianten beschrieben und realisiert wurde (Malone u.a. 1987; Schmid u.a. 1991; Müller 1991b).

Die im Projekt verfolgte Idee eines Informationsmarktes lehnt sich an die Funktion eines gewöhnlichen Warenmarktes an. Der Informationsbedarf im Rahmen verteilter Planungsprozesse wird dem Medium mitgeteilt. Diese Nachfrage löst dort eine Kettenreaktion unter allen möglichen Anbietern (z.B. Datenbanken) aus, bis die Anforderung erfüllt ist oder der Prozeß ergebnislos endet. Das elektronische Medium verfügt über „Anschlüsse" an das soziale System, dem ebenfalls Bedarf mitgeteilt wird. Dieser Leitidee folgt eine softwaretechnische Realisierung, die folgende Merkmale aufweist:

- der Bedarfsdeckungsprozeß geht ohne jede Suche vor sich; er organisiert sich selbst, gesteuert von Angebot- und Nachfrage, von der Sachlogik der Planungsgrößen, nicht von Parametern des Computerprogramms,
- die optimale und vollständige Nutzung aller bekannten Quellen ist immer garantiert,
- mögliche Zielkonflikte werden vollständig aufgedeckt (Konsistenzbeweis),
- der Marktmechanismus ist extrem einfach, zuverlässig und anwendungsunabhängig bei beliebiger Komplexität.

Diese Funktionalität wurde mittels der Technik des Quantitative Reasoning realisiert. Quantitative Reasoning ist ein Verfahren, quantitative, vor allem unscharfe Information aussagenlogisch sinnvoll miteinander zu verknüpfen. Beispiele derartiger Information sind Aussagen wie: Das Unternehmen hat 389 Mitarbeiter; Berlin hat 1289 Einwohner; die Zeitung kostet 1.- DM; die Straße ist 40m breit;

die Firma hat kein Geld; allgemein: Statistiken, Buchhaltungen. Quantitative Aussagen sind immer Ergebnis von Zähl- oder Meßoperationen. Relationale Aussagen lassen sich hieraus durch arithmetische Verknüpfungen, Verhältnisse und Boole'sche Verknüpfungen bilden. Beispiele: Berlin ist doppelt so groß wie Frankfurt; das Durchschnittseinkommen beträgt 2000.- DM/Monat. Der Dollar ist gestern gegenüber dem Yen um 12% gefallen.

Alle wichtigen Fakten und Zusammenhänge eines interessierenden Realitätsausschnitts werden durch Vernetzung solcher Aussagen im Computer repräsentiert. Der Benutzer kann nun in diesem Rahmen Fragen stellen, die das System beantwortet, ohne daß die hierfür notwendigen Schritte programmiert werden müßten. Dieser Prozeß, der die vollständige Funktion des oben beschriebenen elektronischen Informationsmarktes ermöglicht, beruht darauf, daß die (plastisch veränderbaren) unscharfen Zahlen (mit Bandbreiten) sich „auf sich selbst anwenden", indem sie bei Begegnung Eindrücke aufeinander hinterlassen, die in veränderten Werten resultieren. Die Zahlen „kennen" ihre Bedeutung, und daher finden nur sinnvolle Begegnungen und Veränderungen statt. Es handelt sich scheinbar um einen völlig chaotischen Prozeß, bei dem sich alles mit allem kombiniert. Das Ergebnis ist jedoch immer klar nachvollziehbar. Konsistenz oder Inkonsistenz, die der Computer weder beheben kann noch soll, wird hierbei nachgewiesen. Die Einschränkung der Allgemeinheit besteht darin, daß ausschließlich quantitative Aussagen zugelassen sind.

Ein elektronischer Informationsmarkt soll nicht allein das technische Problem der Beschaffung, Kombination, Verdichtung, Verarbeitung und Verteilung sinnvoller Nachrichten durch ein elektronisches Medium lösen, sondern gleichzeitig die Unterstützung räumlich, organisatorisch (d.h. nach Zuständigkeit) verteilt stattfindender Planungsprozesse, also die Kommunikation auf der organisatorischen Ebene, zwischen thematisch unterschiedlich arbeitenden Planern, gewährleisten (Müller 1991b).

Die Grundlagen des Quantitative Reasoning stammen von Beat Schmid (Schmid 1979). Es gibt ein entsprechendes Werkzeug, das aus anwendungsspezifischer Wissensbasis und einer Inferenzkomponente (rechnender und schätzender Teil)

besteht. Die Inferenzkomponente, der einzige zur Laufzeit aktive, anwendungsunabhängige und immer identische Teil des Werkzeugs, besteht aus zwei Teilen. Der eine Teil erkennt die sachlogischen Zusammenhänge und leitet daraus mathematische Beziehungen her, die der zweite Teil zur Berechnung benutzt. Hier handelt es sich um einen Algorithmus, der unter Regelungstechnikern als Kalman-Filter bekannt ist (Kalman 1960).

Auf dieser Grundlage wurden innerhalb des Projekts prototypische Implementierungen vorwiegend für kaufmännische Bereiche (Planung, Controlling) entwickelt. Beispiele: Marktprognosen, Unternehmensberatung, Betriebsdiagnose, Prozeßkostenoptimierung.

6. Zusammenfassung

Das Leitbild der KI hat ungeachtet seiner Kritisierbarkeit Kräfte gebündelt und Energien freigesetzt, die zu Techniken führten, deren Bedeutung auch von KI-Kritikern nicht bezweifelt werden. Diejenigen Unternehmen können sich glücklich schätzen, die in ihrer technischen Kultur Leitbilder mit ähnlicher Durchschlagskraft verankert haben.

Die bis heute geringe Verbreitung von KI-Technologien in Unternehmen liegt (in Übereinstimmung mit Welter (1988)) eher an einer

- Überschätzung der informationstechnischen Dimension der Computeranwendung bei gleichzeitiger
- Unterschätzung (Ignoranz) der sozialen Dimension (wozu auch die Arbeit an differenzierten Leitbildern gehört)

auf Seiten der Designer als an technischen Unzulänglichkeiten. KI- und Software-Krise beruhen „weniger auf mathematisch-logischen oder programmtechnischen Mängeln der bislang verwendeten Methoden des Software-Entwurfs, sondern vielmehr auf der unzureichenden Reflexion des Wechselspiels von technischer Gestaltung und sozialer Wirkung informationstechnischer Systeme" (Coy 1989, S.256).

Appelle zur „stärkeren Einbeziehung" oder „Berücksichtigung" der sozialen Dimension der Informationstechnik blieben in der Vergangenheit ziemlich wirkungslos und werden es auch in Zukunft bleiben, solange sich die Informatik im allgemeinen und die KI im besonderen selbst als rein technische Disziplin verstehen und sich kein adäquateres und tieferes Systemverständnis von der Ausbildung über die Forschung und Entwicklung bis zur Anwendung durchsetzt. Es genügt keineswegs, nicht-technische Aspekte „neben" den technischen zu berücksichtigen. Dies gerät leicht zur folgenlosen Alibi-Beschäftigung von Technikfolgen-Spezialisten. Vielmehr ist bereits auf der metatheoretischen Grundlagenebene der Informatik (und der KI) eine Integration technischer und humanwissenschaftlicher Konzepte erforderlich. Ihr bloßes Nebeneinander mündet nahezu zwangsläufig in die rein moralische Schiene der „menschengerechten Technik".

Mit dem Leitbild des „Computers als Medium" und den darauf aufbauenden softwaretechnischen Konzepten (CSCW, Groupware, VI,...) ist ein wichtiges Etappenziel in Richtung auf die schon lange geforderte Synthese zwischen Informationstechnik und „dem Rest der Welt" erreicht.

Literatur

Applegate, L.; Ellis, C.; Holsapple, C. W.; Radermacher, F. J.; Whinston, A. B. (1991). Organizational Computing: Definition and issues (Editorial). Journal of Organizational Computing, 1, p. 1-10.

Balck, H.; Kreibich, R. (Hrsg.) (1991). Evolutionäre Wege in die Zukunft. Wie lassen sich komplexe Systeme managen? Weinheim u.a.: Beltz.

Bunge, M. (1977, 1979). Treatise on basic philosophy, vol. 3,4 (Ontology I, II), Dordrecht: Reidel.

Clark, D. D. (1991). The Changing nature of computer networks. In: Meyer, A.R. et al. (eds.): Research directions in Computer Science. An MIT perspective. Cambridge, Mass: MIT Press.

Coy, W. (1989). Brauchen wir eine Theorie der Informatik? Informatik-Spektrum, 12, S.256-266.

Coy, W.; Bonsiepen, L. (1989). Erfahrung und Berechnung. Kritik der Expertensystemtechnik. Berlin u.a.: Springer.

Dierkes, M.; Hoffmann, U.; Marz, L. (1992). Leitbild und Technik. Zur Entstehung und Steuerung technischer Innovationen. Berlin: Edition Sigma.

Görz, G. (1991). Künstliche Intelligenz – auf dem Weg zur Anwendung? DEC Seminar, Hamburg, 15. Mai 1991.

Floyd, C. (1989). Softwareentwicklung als Realitätskonstruktion. in Lippe,W. M. (Hrsg.): Software-Entwicklung. Proceedings der Fachtagung, veranstaltet vom Fachausschuß 2.1 der GI, Marburg, 21.-23.6.89; Berlin: Springer.

Haken, H. (1988). Erfolgsgeheimnisse der Natur. Synergetik: Die Lehre vom Zusammenwirken. Frankfurt a. M.: Ullstein.

Haken, H. (1991). Synergetik im Management. In Balck u.a. (Hrsg.). S. 65-91.

Hass, H. (1970). Energon. Das verborgene Gemeinsame. Wien u.a.: Molden.

Kalman,R. E. (1960): A new approach to linear filtering and prediction problems. Journal of Basic Engineering, March 1960, pp 35-45.

Klotz, U. (1991). Die zweite Ära der Informationstechnik. Harvard Manager 2, 1991, S. 101-112.

Luhmann, N. (1988). Soziale Systeme. Frankfurt a. M.: Suhrkamp.

Malone, T. W.; Yates, J.; Benjamin, R. I. (1987). Electronic markets and electronic hierarchies. Comm. ACM 30, 6, S. 484-497.

Malone, T. W.; Rockart, J. F. (1991). Vernetzung und Management. Spektrum der Wissenschaft. November 1991, S. 122- 130.

Malone, T. W.; Crowston, K. (1991). Toward an interdisciplinary theory of coordination. Technical report, Center for Coordination Science, CCS TR# 120, SS WP# 3294-91-MSA, Cambridge, Mass.

Maturana, H. R.; Varela, F. J. (1987). Der Baum der Erkenntnis. Wie wir die Welt durch unsere Wahrnehmung erschaffen – die biologischen Wurzeln des menschlichen Erkennens. Bern u.a.: Scherz.

Müller, R. A. (1988). Systeme verteilter Intelligenz im Unternehmen. Internes Programmpapier der Forschung der Daimler-Benz AG. Berlin.

Müller, R. A. (1991a). Selbstorganisation und Verteilte Intelligenz. Eine Forschungsperspektive der Daimler-Benz AG. In: Balck, H. u.a. (Hrsg.), S. 191-223.

Müller, R. A. (1991b). Kostenplanung mit unscharfen Daten. In Scheer, A.-W. (Hrsg.): Rechnungswesen und EDV. Kritische Erfolgsfaktoren im Rechnungswesen und Controlling. Heidelberg: Physika-Verlag .

Nelius, R. (1992). Entwicklungsstand computergestützter Gruppenarbeit für das Management. Diplomarbeit an der Berufsakademie Stuttgart, angefertigt in der Forschung der Daimler-Benz AG Berlin, Stuttgart.

Probst, G. J. P. (1987). Selbstorganisation. Ordnungsprozesse in sozialen Systemen aus ganzheitlicher Sicht. Berlin u.a.: Parey.

Rammert, W. (1989). Technisierung und Medien in Sozialsystemen. In Weingart, P.: Technik als sozialer Prozeß. Frankfurt am Main: Suhrkamp, S. 128-173.

Rammert, W. (1992). „Expertensysteme" im Urteil von Experten. In: Jahrbuch 6, „Technik und Gesellschaft" Frankfurt a.M.: Campus.

Schmid, B. (1979). Bilanzmodelle. Simulationsverfahren zur Verarbeitung unscharfer Teilinformationen. Bericht des ORL-Instituts der ETH Nr. 40, Zürich.

Schmid, B. u.a. (1991). Die elektronische Revolution der Märkte. IO Management Zeitschrift, 60, Heft 12, S. 96-98.

Staehle, W. H. (1987). Management. München.

Trist, E. L. (1981). The Socio-technical perspective – The evolution of socio-technical systems as a conceptual framework and as an action research program. In: Ven, A. van der; Joyce, W. F. (Eds.): Perspectives on organization design and behavior. New York.

Turoff, M. (1991). Computer-mediated communication requirements for group support. Journal for Organizational Computing, 1 (1), S. 85-113.

Ulrich, H.; Probst, G. (1990). Anleitung zum ganzheitlichen Denken und Handeln. Bern.

Varela, F. J. (1990). Kognitionswissenschaft – Kognitionstechnik. Frankfurt a. M.: Suhrkamp.

Welter, G. (1988). Technisierung von Information und Kommunikation in Organisationen. Eine kritische Analyse der Entwicklung und des Einsatzes informations- und kommunikationstechnischer Systeme.Spardorf: Wilfe.

Winograd, T.; Flores, F. (1986). Understanding computers and cognition. A new foundation for design. Reading, Mass.: Addison Wesley.

Womack, J. P.; Jones, D. T.; Ross, D. (1991). Die zweite Revolution in der Autoindustrie. Frankfurt: Campus.

KI – Perspektiven der Anwendung und Technologiefolgenabschätzung

R. Haberbeck

„Zusammenfassend ist zu sagen, daß der Grundirrtum, der jede Erwägung unfruchtbar macht, darin besteht, die Technik als ein in sich abgeschlossenes Kausalsystem zu sehen. Dieser Irrtum führt zu jenen Unendlichkeitsphantasien, in denen sich die Begrenzung des reinen Verstandes verrät. Die Beschäftigung mit der Technik wird erst dort lohnend, wo man sie als Symbol einer übergeordneten Macht erkennt." Ernst Jünger

„Die Technik entwickelt sich immer vom Primitiven über das Komplizierte zum Einfachen", Antoine de Saint-Exupéry

Einleitung

Dieser Beitrag betrachtet die Orientierung der Künstlichen Intelligenz am Verwendungszusammenhang. Dieser lösungsorientierte Ansatz zeigt, daß es nicht eine spezifische Technologiefolgenabschätzung für die Künstliche Intelligenz geben muß. Konzeptquellen der Künstlichen Intelligenz – auch die, die nicht auf dem Verwendungszusammenhang basieren, sondern auf abstrahierten Aspekten der menschlichen Intelligenz– lassen die Anwendung der Künstlichen Intelligenz in dem Anwendungsumfeld aufgehen. Die Trennung von Künstlicher Intelligenz und traditioneller Informationsverarbeitung hinsichtlich der Technologiefolgenabschätzung fällt schwer, was auch allgemein für die Unterscheidung zwischen Künstlicher Intelligenz und traditioneller Informationsverarbeitung gilt. Leitvorstellungen und Konzeptquellen, die sich am Verwendungszusammenhang orientieren, bieten die Möglichkeit, Technologiefolgenabschätzung schon von vornherein in der Konzeption der anwendungsorientierten Entwicklung der Künstlichen Intelligenz zu betreiben. Technologiefolgenabschätzung und Leitvorstellungen der Künstlichen Intelligenz und somit deren Verantwortbarkeit ist ein Problem des umfassenden Managements und der frühen Verwendungsorien-

tierung (Benutzer- und Marktorientierung) der Konzeption und der Entwicklung der Künstlichen Intelligenz.

Die Orientierung der Künstlichen Intelligenz am Verwendungszusammenhang

Die Auffassung der Künstlichen Intelligenz im weiten Sinn (Orientierung der maschinellen Informationsverarbeitung am Menschen) beinhaltet die Gestaltung von Informationstechnologien, die an dem Verwendungszusammenhang der angestrebten Lösung orientiert ist. Hierbei geht es nicht um das Simulieren von isolierten Aspekten der menschlichen Intelligenz wie z.B. Spracherkennen, visuelles Erkennen oder Entscheiden, sondern es geht um die Gestaltung der Interaktion von Mensch und Computer (oder um die Nutzung des Computers durch den Menschen), die durch die Interaktionsweise des Menschen (oder am Menschen) im integrierten Verwendungszusammenhang der Informationstechnologie orientiert ist.

Die Konzeption der Künstlichen Intelligenz in diesem Sinn orientiert sich an konkreten Problemen, die in der Mensch Maschine Interaktion oder anderen Problemstellungen im Einsatz von Computern zu verbessern oder neu zu gestalten sind. Es werden hierbei verschiedene andere Konzeptquellen der Künstlichen Intelligenz unter Umständen kombiniert. Die Leitvorstellung (die Konzeptquelle), die die Entwicklung der Künstlichen Intelligenz in diesem Umfeld bestimmt, ist die integrierte, anwendungsorientierte Lösung, nicht abstrahierte Aspekte der Intelligenz des Menschen. Diese verwendungsorientierte Konzeptquelle der Künstlichen Intelligenz steht auch der folgenden, allgemeinen Definition von Künstlicher Intelligenz nahe: „KI läßt sich als Nachbildung natürlicher Intelligenz dann realisieren, wenn definierbare und objektivierbare Teilfunktionen von Intelligenz auf maschinelle Systeme übertragen werden und wenn dabei introspektive Aspekte wie das Verstehen von Inhalten keine Rolle spielen. Man kann derartige Teilaspekte von Intelligenz als operationale Intelligenz bezeichnen oder auch – in Abgrenzung zu umfassender menschlicher Intelligenz – als Pseudo-Intelligenz" (Voigt 1991).

Vom Standpunkt des Entwicklers oder Ingenieurs, der eine Programmiersprache verwendet, die dem Paradigma der Künstlichen Intelligenz zuzurechnen ist (z.B. LISP oder Prolog), ist der Einsatz solch einer Programmiersprache und einer entsprechenden Entwicklungsumgebung schon ein Indiz, daß die Entwicklung, die er betreibt, als Künstliche Intelligenz aufzufassen ist. Die in diesem Fall behandelte Lösung muß nicht unbedingt als typisch im Rahmen der Künstlichen Intelligenz vom Standpunkt des Anwenders verstanden werden.

Andererseits kann ein Anwender eine Lösung als ein Beispiel für die Anwendung von Künstlicher Intelligenz auffassen. Dieses Programm muß jedoch vom Entwickler oder Ingenieur nicht als ein Beispiel für Künstliche Intelligenz gesehen werden. Es wurde nicht in einer typischen Programmiersprache der Künstlichen Intelligenz und nicht mit Hilfe einer entsprechenden Entwicklungsumgebung erstellt. Fortran oder C und ein einfacher Editor als Entwicklungsumgebung lassen die Programmerstellung vom Standpunkt des Entwicklers nicht als Künstliche Intelligenz erscheinen.

Der Unterschied in den Standpunkten des Entwicklers und des Anwenders, was als Künstliche Intelligenz zu verstehen ist, ist fundamental. Die Entscheidung, ob eine Entwicklung oder Anwendung als Künstliche Intelligenz aufzufassen ist, hängt vom Betrachtungs- und Bewertungsstandpunkt ab. Eine Entwicklung oder Anwendung kann nicht von vornherein und voraussetzungslos als Künstliche Intelligenz oder als nicht zu diesem Paradigma zugehörig bewertet werden. Selbst die Einbeziehung der Relativität dieses Urteils bezüglich fundamentaler Standpunkte läßt nicht ein eindeutiges Urteil zu (Künstliche Intelligenz oder nicht). Eine skalierte Bewertung erscheint sinnvoll, wo abhängig vom Standpunkt von „x %" Künstlicher Intelligenz in der Anwendung oder Entwicklung gesprochen werden könnte. Wünschenswert ist ein verwendungsfähiges, allgemein anerkanntes und standpunktbezogenes Konzept von Künstlicher Intelligenz, das auch eine entsprechende Skala einschließt, die solch ein abgestuftes Urteil zuläßt. "The question 'Is it AI?' highlights a misconception. You can write ordinary programs in Lisp, supposedly an AI language. And you can write AI in C, a procedural language. Rule-based systems give you a lot of flexibility, while more traditional programming systems give you better speed. It's really a con-

tinuous spectrum. In the course of development, a project can shift back and forth across that spectrum. What's important is incorporating feedback from experts and users into the development cycle." (Heller 1991)

Schlachetzki (in diesem Band) zeigt auf, daß die physikalische Gestalt des Programmablaufs (binäre/diskrete Elektronen- oder Photonenrechner oder Analogrechner) immer auf der Basis eines binären/diskreten, deterministischen Modells geschieht – auch die Verwendung selbstadaptierender neuronaler Netze oder sogenannter „Fuzzy Logic". Die Unterschiede vom Standpunkt des Entwicklers hinsichtlich der Verwendung von Programmiersprachen und Entwicklungsumgebungen aus dem Bereich der Künstlichen Intelligenz oder aus dem Bereich des traditionellen Programmierens erscheinen unter diesem Blickwinkel sinnlos. Die physikalische Wirklichkeit des Programmablaufs nivelliert die Unterschiede der Programmiersprachen (Künstliche Intelligenz oder traditionelle Programmiersprachen) vom Standpunkt des Entwicklers. Auch die Sichtweise des Anwenders, ob bei einer Anwendung „x %" Künstliche Intelligenz vorliegt, wird bei der Betrachtung vom Standpunkt der Gestalt des physikalischen Programmablaufs hinfällig.

In diesem Zusammenhang ist auf die Bedeutung der mathematischen Logik – der „Mutter aller exakten Darstellungsformen" – hinzuweisen, die auf einem binären/diskreten und deterministischen Konzept basiert. Cremers/Eder/Hinze (in diesem Band) verweisen auf den fundamentalen Charakter der mathematischen Logik für die Künstliche Intelligenz auf der Ebene der Wissensrepräsentation und des Schließens – aber auch auf die Grenzen klassicher Logik. Fundamentale Konzepte der Künstlichen Intelligenz und die Gestalt des physikalischen Programmablaufs erscheinen identisch.

Einige Beispiele sollen verdeutlichen, daß Künstliche Intelligenz in der Orientierung am Verwendungszusammenhang mittlerweile häufig vorzufinden ist, aber nicht unbedingt als Anwendung der Künstlichen Intelligenz von den Anwendern Zusatz Anfang oder Entwicklern verstanden wird.

Ein Beispiel ist die Verwendung von Orthographie-Hilfen (spell checker), die heute in fast jedem Textverarbeitungsprogramm enthalten sind. Die Kompetenz eines Spezialisten für Rechtschreibung ist selbstverständlicher Teil einer Texterstellungsanwendung. Die intelligente Leistung eines Texterstellungsprogramms – die Rechtschreibprüfung – ist ein Aspekt der Gesamtlösung. Kaum ein Anwender eines Textverarbeitungsprogramms wird die Rechtschreibehilfe in der Nutzung des Texterstellungsprogramms als besondere Leistung der Künstlichen Intelligenz wahrnehmen.

Es stehen heute Programme für das Komponieren und Erstellen von Musik für verschiedenste Stilrichtungen auf Personal Computern für den professionellen Bereich und für Freizeitanwendungen auf dem Markt zur Verfügung, die die Leistungen guter Komponisten oder Musiker erreichen. Diese Programme sind für einige hundert Mark zu erwerben und können schnell von musikalischen Laien erlernt und bedient werden. Diese Programme werden nicht als Künstliche Intelligenz von den Herstellern angeboten und sie werden auch nicht als Künstliche Intelligenz von den Anwendern aufgefaßt, obwohl diese Produkte Eigenschaften aufweisen, die sie vom Standpunkt des Entwicklers als Künstliche Intelligenz einordnen lassen. Die Fähigkeiten und Erfahrungen hochqualifizierter Musiker und Komponisten sind ohne ein tiefes Musikverständnis des Anwenders von einem preisgünstigen Personal Computer (Heimcomputer) abrufbar und erscheinen dem Anwender als praktisch und unterhaltsam – aber nicht als Künstliche Intelligenz.

Ein weiteres Beispiel in diesem Sinn ist die Bedienung eines Computers mit der Maus und einer entsprechenden graphischen, objekt-orientierten Benutzeroberfläche. In diesem Fall wird die für die menschliche Interaktion elementare Zeigefunktion – z.B. Zeigen mit dem Finger in Zusammenhang mit einer kontextabhängigen Bedeutung und eventuell einer Handlungsaufforderung – in der Mensch Maschine Interaktion verwendet. Vor 15 Jahren erschien die Verwendung der Zeigefunktion in dem beschriebenen Zusammenhang als innovativ. Heute ist eine solche Anwendung kaum der Rede wert.

Technologiefolgenabschätzung und Künstliche Intelligenz

Die Technologiefolgenabschätzung der Künstlichen Intelligenz im Rahmen dieser am Verwendungszusammenhang orientierten Konzeptquelle bringt unmittelbar mit sich, daß eine scharfe Trennung der Technologiefolgenabschätzung der Künstlichen Intelligenz von der Informationsverarbeitung schlechthin nicht zu vollziehen ist. Künstliche Intelligenz in dem oben beschriebenen Sinn ist von vornherein in Informationstechnologien eingebettet, die nicht notwendig als isolierte Anwendungen der Künstlichen Intelligenz zu verstehen sind. Es hat sich im Laufe der Entwicklung der Künstlichen Intelligenz gezeigt, daß das reine Spracherkennen (die Transformation von akustisch-phonetischen Ereignissen in geschriebene Texte) erfolglos ist ohne die Einbeziehung der Bedeutung des Gesagten (der syntaktisch-semantischen Interpretation) und ohne Beziehung zum Kommunikationskontext der Kommunikationsteilnehmer (semantisch-pragmatische Interpretation, Hintergrundwissen).

Die Prognosen von renommierten Beratungsfirmen zur Marktentwicklung eines Gebietes der Künstlichen Intelligenz – der Sprachverarbeitung – wiesen Mitte der 80er Jahre auf einen zweistelligen Milliarden Dollar Markt für den Anfang der 90er Jahre hin. Diese Prognosen erwiesen sich – wie viele Prognosen zum technischen Durchbruch oder zum Markterfolg der Künstlichen Intelligenz in den Jahrzehnten davor – als falsch. Die Schwierigkeit, technische Trends oder die Marktentwicklung für den Bereich der Künstlichen Intelligenz vorherzusagen, macht deutlich, wie schwierig die Technologiefolgenabschätzung in diesem Bereich ist. Der Bezug auf reine technologische Trends und Entwicklungsdimensionen sowie die Vernachlässigung der Anwender in frühen Entwicklungsphasen können solche falschen Prognosen erklären. Müller (in diesem Band) weist auf die Einbettung der Künstlichen Intelligenz in traditionelle Informationsverarbeitungssysteme hin. Damit zusammenhängende Prognosen und entsprechende Technologiefolgeabschätzungen, die spezifisch für die Künstliche Intelligenz sind, sind nahezu unmöglich. "For now, AI quietly continues to migrate from development groups into everyday applications in hopes of living up to its much ballyhooed past" (Francett 1991).

Wie in der Verarbeitung von Sprache in der Künstlichen Intelligenz verschiedene Ansätze in der Vergangenheit sich zu integrierten Herangehensweisen entwickelten, so wird dieser Trend generell die Zukunft der Künstlichen Intelligenz prägen, die sich mit abstrahierten Aspekten der menschlichen Intelligenz befaßt. Dies bedeutet jedoch nicht, daß z.B. die gesprochene Sprache das optimale oder einzige Kommunikationsmittel zwischen Mensch und Computer sein wird. Untersuchungen der Nixdorf Computer AG Ende der 80er Jahre zeigten, daß der Einsatz von Spracherkennung beim professionellen Bedienen von integrierten Büroprogrammen (Textverarbeitung, Terminverwaltung) keinen Vorteil in der Bedienungsgeschwindigkeit und Bedienungssicherheit bietet. Die Bedienung mit Tastatur und Maus erwies sich unter diesem Blickwinkel sogar als vorteilhaft. Eine frühe Einbeziehung von Anwendern in den Entwicklungszyklus erwies sich als sinnvoll für die Entwicklung und Produktgestaltung und ersparte technologische Irrwege und Sackgassen, wie sie in einigen Studien von renommierten Beratungsfirmen prognostiziert wurden. Die Gestaltung und Bewertung der Mensch Maschine Interaktion mit Anwendern wird in der Siemens Nixdorf Informationssysteme AG auch in anderen Bereichen (z.B. Entwicklung von Selbstbedienungsterminals für verschiedene Bereiche) in frühen Phasen der Projektentwicklung eingesetzt (Haberbeck 1991).

Der durch einen Fehler des Piloten, der die Anweisungen des Bordcomputers mißachtete, verursachte Absturz eines Airbus ließ das Vertrauen der Passagiere in die Zuverlässigkeit dieses Flugzeugtyps nicht sinken. Benutzer von Flugzeugen sind – unter Umständen – indirekt Anwender von Künstlicher Intelligenz (abhängig vom Standpunkt). Bei der Frage nach der Zuverlässigkeit des Airbus wurde das Thema Künstliche Intelligenz in der Öffentlichkeit nicht diskutiert. Vom technischen Standpunkt ist der Bordcomputer als Anwendung Künstlicher Intelligenz anzusehen, was aber für die öffentliche Diskussion der Anwender unwichtig ist. Die Einbettung in ein komplexes System läßt den vom technischen Standpunkt erwogenen Aspekt der Künstlichen Intelligenz verschwinden.

Der Zusammenbruch des Börsenhandels 1988 ist auf unzureichend programmierte Computer der Börsenhändler zurückzuführen. Dies wurde von den Anwendern und in der Öffentlichkeit nicht als ein Versagen der Anwendungen

der Künstlichen Intelligenz gesehen, obwohl vom technischen Standpunkt aus die verwendeten Programme als Künstliche Intelligenz zu bezeichnen sind. Das System Börse brach zusammen – was auch ohne die Verwendung von Computern passieren kann – in Zusammenhang mit unzureichend programmierten Computern. Der in Erwägung zu ziehende Aspekt Künstliche Intelligenz spielt dabei keine Rolle. Die optimierten Programme – so die Betroffenen – werden hoffentlich in Zukunft besser arbeiten.

Die oben angeführten Beispiele machen deutlich, daß die Einbettung der Künstlichen Intelligenz in den Verwendungszusammenhang eine für die Künstliche Intelligenz spezifische Technologiefolgenabschätzung erschwert. Neben den Kategorien wie Erleichterung der Benutzung, einfacheres Erlernen der Anwendung, die die positiven Aspekte der Technologiefolgenabschätzung ansprechen, sind auch Kategorien der negativen Aspekte der Technologiefolgenabschätzung zu berücksichtigen wie unabsehbare Nebeneffekte im Masseneinsatz, Langzeitfolgen.

Auch wenn Konzeptquellen der Künstlichen Intelligenz nicht von vornherein den Verwendungszusammenhang der Anwendung der Technologie bedenken, so ist jedoch davon auszugehen, daß reine, isolierte Anwendungen von Künstlicher Intelligenz nicht in der Wirklichkeit der Welt der Anwendungen vorzufinden sind oder sein werden. Auch der Einsatz des Übersetzungsprogramms METAL® der Siemens Nixdorf Informationssysteme AG, dessen nahezu zwanzigjährige Entwicklung von Anfang an anwenderorientiert erfolgte mit der Einbeziehung von Benutzern, wird von den heutigen Kunden nicht als spezifische Anwendung der Künstlichen Intelligenz wahrgenommen. Entsprechend sind auch das Marketing und der Vertrieb gestaltet.

Perspektiven der Technologiefolgenabschätzung und der Leitvorstellungen der Künstlichen Intelligenz

In der Technologiefolgenabschätzung der Informationstechnologie wird sich die Künstliche Intelligenz nicht besonders hervorheben. Die bisher vorzufindenden

Anwendungen der Künstlichen Intelligenz im weiten Sinn machen deutlich, daß die Einbettung in den Verwendungszusammenhang – ob intendiert oder nicht – nicht zu der Entwicklung einer spezifischen Technologiefolgenabschätzung führen muß.

Der Einsatz von Verfahren der Künstlichen Intelligenz wird immer stärker Bestandteil von professionellen Anwendungen wie auch bei Anwendungen der Informationstechnologie im Freizeitbereich. Computerspiele sind immer mehr als Anwendungen von Künstlicher Intelligenz zu verstehen. Programme zur Komposition von Musik auf Heimcomputern sind den Kompositionsfähigkeiten durchschnittlicher menschlicher Komponisten/Musiker ebenbürtig, wenn nicht sogar überlegen. Haushaltsgeräte werden durch Fuzzy Logic gesteuert. All diese Phänomene nimmt der Anwender nicht als Anwendungen der Künstlichen Intelligenz wahr, sondern als Bereicherung seiner Lebensqualität.

Solange der Einsatz von Künstlicher Intelligenz in einen Verwendungszusammenhang eingebettet ist, bestimmt dieser Verwendungszusammenhang die Leitvorstellung der Künstlichen Intelligenz. Dies kann dazu führen, daß standpunktabhängig unter der Dominanz der Leitvorstellung des Verwendungszusammenhangs die Künstliche Intelligenz nicht wahrgenommen wird oder nicht zur Geltung kommt. Wird Künstliche Intelligenz von vornherein für Verwendungszusammenhänge konzipiert und orientiert sich nicht an abstrahierten Aspekten der menschlichen Intelligenz, so ist durch die Anwendungsorientierung der Leitvorstellung auch der Erfolg der Anwendung der Künstlichen Intelligenz mitbestimmt. Isolierte Aspekte der Intelligenz des Menschen zu Leitvorstellungen für ein Paradigma wie die Künstliche Intelligenz zu nehmen, kann entweder zu nicht erfolgreichen Entwicklungen führen oder wirft die Frage nach einer Technologiefolgenabschätzung auf, die spezifisch für die Künstliche Intelligenz sein müßte.

In dem letzten Fall ist für die Technologiefolgenabschätzung für die Künstliche Intelligenz aktuell kein Gegenstand im Bereich der marktfähigen Anwendungen auszumachen. Sollte sich jedoch in Zukunft eine Konzeption von Künstlicher Intelligenz durchsetzen, die nicht auf Aspekte der menschlichen Intelligenz ein-

geschränkt ist (also keine Pseudo-Intelligenz), so ist eine besondere Sorgfalt auf die Gestaltung der Anwendung, der technischen Entwicklung und der zugehörigen Leitvorstellungen zu legen im Sinne einer präventiven Technologiefolgenabschätzung. Eine Orientierung der Leitvorstellungen an dem technisch Machbaren oder ähnlich naiven oder abstrakten Konzepten unter Vernachlässigung des Verwendungszusammenhangs läßt einen Mangel an Verantwortbarkeit deutlich werden, der die Entscheidungsautonomie des Menschen in Zukunft gefährden könnte.

Zukünftige Aktivitäten in diesem Gebiet sollten die Aufgabe und Überwindung der abstrakten und technisch-wissenschaftlichen Konzeptquellen und Leitvorstellungen thematisieren zum Vorteil einer integrierten, anwendungsorientierten Gestaltung von Konzeptquellen und Leitbildern der Künstlichen Intelligenz (wie auch immer der Name für eine Leitvorstellung oder Konzeptquelle heißen wird). Ein Ansatz in dieser Hinsicht ist das „real-World Computing Program" der japanischen Regierung, wobei eine Marktorientierung bei diesem Programm nicht deutlich auszumachen ist. Die Einbeziehung von Anwendern in die frühe Phase der Entwicklung bietet die Möglichkeit einer markt- und anwendungsorientierten Entwicklung, die das Risiko von Fehlentwicklungen und Fehlprognosen gering hält. Das Vermeiden von technologischen Irrwegen und Sackgassen wird sich auch positiv auf die Kosten der Entwicklung und somit auf die Preise der Produkte auswirken. Zukünftige Programme oder Projekte zur Technologiefolgenabschätzung im Gebiet Künstliche Intelligenz sollten unter Einbeziehung von Industrie, Forschungseinrichtungen, Staat und einer entsprechenden Öffentlichkeit von Benutzern die Verantwortbarkeit entsprechender Anwendungen gewährleisten. Dies zielt nicht auf eine Verlangsamung und Verteuerung des Entwicklungs- und Produktzyklus, sondern eine integrierte, anwendungsorientierte Beschleunigung und Optimierung ist das Ziel.

Literatur:

Francett, B. (1991). AI (quietly) goes mainstream. Computerworld, Juli 1991.

Haberbeck, R. (1991). M4-EVAL: Evaluation of multi-medial and multi-modal human-computer interaction. In: H.-J. Bullinger (Hg.): Human aspects in computing Amsterdam: Elsevier, S. 1285-1294.

Heller, Martin (1991). AI in practice. BYTE, Januar 1991.

Voigt, Hartmut von (1991). Künstlich oder intelligent. Dialog 2, 1991, S. 10-15.

Abschnitt III:

Zukunftsauswirkungen der Künstlichen Intelligenz

Einführung und Übersicht zum Abschnitt III

I. Wachsmuth, M. Wilker

Die theoretische Künstliche-Intelligenz-Forschung hat sich die folgende Aufgabe gestellt: Jeder Aspekt menschlicher Intelligenz soll so präzise beschrieben werden, daß er durch ein künstliches System – einen Computer – simuliert werden kann. Daß dies prinzipiell möglich sei, bezieht sich auf eine Ausgangshypothese, nämlich daß Menschen – zumindest als intelligent Handelnde – informationsverarbeitende Systeme sind. Verband sich mit den Ursprüngen der KI (Dartmouth Conference 1956) in erster Linie der Gedanke, auf diese Weise Erklärungsmodelle für Intelligenz zu erhalten, so wird bei Anhängern der sog. „harten KI" die Auffassung verteten, daß es keine prinzipiellen Unterschiede zwischen menschlichen Denkleistungen und ihrer Nachbildung auf Maschinen gibt. Das hieße, daß sich menschliche Intelligenz auf dem Computer reproduzieren läßt und – von der anderen Seite her betrachtet – daß menschliche Intelligenz sich auf Informationsverarbeitung reduzieren läßt.

Die Künstliche-Intelligenz-Forschung befaßt sich aber auch mit der Konstruktion von Systemen, die kognitive Leistungen erbringen, um die theoretisch entwickelten Konzepte und Techniken nutzbringend einzusetzen: Auf informationsverarbeitende Maschinen werden Eigenschaften menschlicher Intelligenz, etwa Schlußfolgerungsfähigkeiten, übertragen; dadurch sollen geistige Tätigkeiten des Menschen unterstützt, verstärkt oder entlastet werden. In dieser häufig als „weiche KI" charakterisierten Richtung beschränkt sich die Erwartung der erzielbaren Leistungen auf Systeme, die in abgrenzbaren Bereichen als Werkzeuge menschliche Intelligenzfähigkeiten substituieren oder sie möglicherweise auch weit übertreffen, die aber nicht menschengleich sein werden. Selbst wenn die Realisierung einer intelligenten Maschine unerreicht ist und bleibt, beginnt die Forschung über „Künstliche Intelligenz" zum Verständnis des Denkens und des Menschen beizutragen.

Durch den in der Künstlichen Intelligenz vertretenen Anspruch, *geistige* Tätigkeiten des Menschen zu formalisieren und auf Maschinen zu übertragen, wird eine neue Qualität technischer Mittel berührt. Die mit einer „denkenden Maschine" verbundenen Möglichkeiten und utopischen Vorstellungen gehen einher mit Umgestaltungen der Arbeits- und Lebenswelt. Sie berühren die Neubewertung beruflicher Qualifikationen, die Veränderung von Werten, ja selbst des *Menschenbildes*. Die Implikationen solcher Vorstellungen werden in diesem Buchabschnitt diskutiert. Nach Ansicht der Autoren haben die beiden als „harte" bzw. „weiche" KI bezeichneten Standpunkte völlig verschiedene Menschenbilder zum Gegenstand, die die Art und Weise, wie Künstliche Intelligenz in gesellschaftliche Wechselwirkungen tritt, betreffen und damit die Entwicklung der zukünftigen informationstechnischen Gesellschaft fundamental beeinflussen könnten. So wird gegenwärtig nicht ausschließlich davon gesprochen, inwieweit „intelligente" Funktionen in technischen Systemen realisiert werden, sondern auch davon, daß maschinelle „Akteure" als Kooperationspartner in menschliche Arbeitszusammenhänge eintreten.

Die in den Medien dazu diskutierten Fragen schwanken zwischen euphorischen Zukunftsvisionen einerseits und greifen andererseits menschliche Urängste vor künstlichen Kreaturen auf. Der mit den gegenwärtig begrenzten Erfolgen und möglicherweise inhärenten Schwächen solcher Systeme Vertraute vermag sich schwerlich vorzustellen, daß in absehbarer Zeit eine Konkurrenz zum intelligent handelnden Menschen in seiner Gesamtheit erwachsen könnte. Eine denkende Maschine, die sich ihrer selbst bewußt ist, bleibt vielleicht immer dem Bereich der Science Fiction zuzurechnen. Aber selbst dann äußert sich in den geführten Debatten das offensichtliche Bedürfnis, sich mit einer solchen Zukunftsvision zu beschäftigen, bevor sie auch nur ansatzweise Realität wird. Aus den angesprochenen Gründen muß ein solcher Diskurs Fragen des Menschenbildes einschließen.

Zwar gibt es kein konsensfähiges, universell gültiges Menschenbild; aber es gibt Menschenbilder, die in bestimmten Epochen und bestimmten Regionen eindeutig dominieren und darum relevant sind. In den hochentwickelten Industriegesellschaften war in den letzten Jahrzehnten (innerhalb und außerhalb des

Wissenschaftssystems) ein Menschen- und Weltbild erfolgreich, das den Menschen als zwar hochkomplexe, aber durchgängig determinierte und kausalen Gesetzmäßigkeiten unterworfene Maschine darstellt. Die „Maschine" Mensch, wie der sie umgebende Kosmos, ist prinzipiell verstehbar und deshalb symbolisch rekonstruierbar. Dieses Axiom ist die Grundvoraussetzung des Postulats einer „harten" KI. Das menschliche Hirn gilt als Sitz des Bewußtseins, welches als die Summe der im Hirn ablaufenden biophysikalischen Prozesse aufgefaßt wird.

Schon immer ist ein Menschenbild kritisiert worden, das menschliches Handeln auf kausale Prozesse reduziert. Innerhalb des Wissenschaftssystems wird dies in jüngerer Zeit vor allem in der Chaostheorie bestritten: Kausalität ist nur ein Ausschnitt aus einer umfassenderen Wirklichkeit; für die „harte" KI könnte das bedeuten, daß Bewußtsein sich als ein Phänomen erweist, das zwar von kausalen Prozessen begleitet wird, aber nicht aus ihnen besteht. Außerhalb des Wissenschaftssystems manifestiert sich die Vorstellung einer vollständig determinierten Welt in einer Ideologie der Machbarkeit, die von einer ökologischen Krise begleitet und darum zunehmend angefochten wird.

Inwiefern kann eine *entwickelte* Künstliche Intelligenz, welche (gegenwärtig nicht erreichte) Zielvorstellungen der anfangs angesprochenen Art realisiert, Auswirkungen auf die konkurrierenden Menschenbilder haben? Es könnte sich herausstellen, daß es trotz aller Anstrengungen unmöglich ist, eine bewußtseinsfähige Maschine zu bauen. Sollte sich dabei zeigen, daß die Unmöglichkeit nicht auf einem Komplexitätsgefälle zwischen hirnorganischen Vorgängen und ihren Simulationen beruht, sondern prinzipieller Natur ist, bedeutete das das Ende des naturwissenschaftlichen Instrumentalismus und damit das Ende einer geistesgeschichtlichen Epoche.

Außerwissenschaftliche, „vulgäre" Menschenbilder sind derzeit geprägt von einem anthropologischen Pessimismus, der weit über eine Fin-de-siècle-Stimmung hinausgeht. In zunehmendem Maße begreifen sich die Menschen in den hochentwickelten Industrieländern als Wesen, die durch ihre technologischen Fähigkeiten ihre ökologischen und sozialen Voraussetzungen untergraben. Her-

kömmliche Technologien werden zunehmend als Beschleuniger ökologischer und sozialer Entropie wahrgenommen. Ein solches pessimistisches Menschenbild könnte korrigiert werden, sofern sich herausstellt, daß eine entwickelte KI einen Beitrag zur Lösung globaler Probleme leisten kann, etwa indem sie die Ressourcenverteilung optimiert oder durch neuartige Informationssysteme den Verkehrskollaps verhindert.

Künstliche Intelligenz stellt eine besondere Herausforderung für die Zukunftsgestaltung dar: Auch wenn die Realisierung „entwickelter" KI noch nicht absehbar ist, soll man sich daher schon jetzt mit ihren möglichen Auswirkungen befassen. *Zukunft sollte keine passive Angelegenheit sein, in die man sich treiben läßt, sondern es ist uns aufgegeben, die Zukunft aktiv zu gestalten. Die Richtungen, in die sich die Zukunft durch Künstliche Intelligenz entwickeln könnte, sind nicht zufällig, sondern sie werden geleitet von den mehr oder weniger expliziten Spekulationen der Beteiligten oder ihrer Auftraggeber.*

So spekuliert etwa der amerikanische Computerwissenschaftler Hans Moravec darauf, daß – unter der Voraussetzung, daß Computer zehn Billionen Rechenoperationen pro Sekunde schaffen werden – in etwa 50 Jahren menschliche Intelligenz inklusive ihrer intuitiven Fähigkeiten auf einem Rechner erzeugt werden kann. Seine Zukunftsvision beinhaltet drastische Veränderungen der gesellschaftlichen Verhältnisse und schließliche Ablösung des Menschen durch Maschinen: „Die Maschinen können jede Arbeit verrichten. Sie werden die Ingenieure ersetzen. Die Unternehmen werden von Management-Systemen geleitet, die wiederum von Computern kontrolliert werden.[...] [Die Menschen] werden die Verbraucher sein. Ich glaube, auf lange Sicht sind die Menschen für das Funktionieren und die Entwicklung der Gesellschaft irrelevant. [...] Man kann [die Maschinen] aber auch als unsere Produkte sehen, Erzeugnisse, die eines Tages, wenn sie selbständig sind, unsere Kultur erben sollen."[1]

[1] Auszüge aus einem Interview von Helene Conrady mit Hans Moravec (Abdruck in den VDI-Nachrichten vom 6. Dezember 1991).

Die technische Realisierbarkeit einmal beiseite gelassen, steht hier vor allem die Wünschbarkeit solcher Entwicklungen zur Debatte, die eine wohl denkbar drastische Änderung heutiger Vorstellungen vom Menschen bedeuten würde. Auch wenn die Machbarkeit von Moravecs Spekulationen gegenwärtig kaum ernsthafte wissenschaftliche Diskussion findet, darf doch ihr Einfluß auf mögliche Zukunftsentwicklungen nicht unterschätzt werden. So haben Beispiele aus jüngerer Zeit – zu denken ist etwa an SDI und Robotik – gezeigt, daß Visionen des bisher kaum Vorstellbaren Leitlinie für die Mobilisierung von Forschungsprogrammen und -investitionen sein können, deren Umsetzung auch bei nur teilweisem Erreichen der Ziele kaum zurücknehmbare Auswirkungen nach sich zieht.

Die Beiträge dieses Abschnitts entstanden innerhalb eines Diskurses unter Teilnehmern unterschiedlicher Denktraditionen und Arbeitsrichtungen, der um die möglichen Auswirkungen einer – hypothetisch betrachteten – entwickelten KI geführt wurde. Es konnte erwartungsgemäß nicht gelingen, zu einer auch nur ansatzweise zufriedenstellenden Beantwortung der im Verlauf des Diskurses aufkommenden Fragen zu gelangen. Daher liegt der Wert des Erreichten eher in der Entwicklung von Kategoriengefügen, vor deren Hintergrund eine Behandlung der durchweg als drängend empfundenen Zukunftsfragen in reflektierter Weise möglich ist, als in deren letzt?gültiger Beantwortung – was ein wohl illusorisches Ziel darstellen würde.

Im folgenden Abschnitt setzt sich A. Kremeier mit der Rolle der Informationstechnik im gesellschaftlichen System auseinander und stellt damit den allgemeinen Hintergrund für gesellschaftliche Wechselwirkungen der KI bereit. Der daran anschließende Gruppenbeitrag der Teilnehmer des Zukunftsdiskurses betrachtet mögliche Auswirkungen einer entwickelten KI entlang gestaffelter Stadien, die hypothetisch Merkmale zukünftiger KI-Systeme beschreiben. Fragen der Verantwortung von KI-Produkten greift P. Schreiber auf, die sich auch mit neuen Rollenverteilungen zwischen Mensch und Maschine auseinandersetzen, und H. Röpke behandelt schließlich potentielle Gefahren von KI-Systemen.

Informationstechnik im gesellschaftlichen System

A. Kremeier

Einleitung

Künstliche Intelligenz umfaßt sowohl eine Disziplin, die Theorien über Beschaffenheit und Funktion von Geist, Denken, Wissen und Intelligenz aufstellt, wie einen Zweig der technischen Wissenschaften. Er entwirft Maschinen, denen geistige Tätigkeiten des Menschen übertragen werden können (Krämer 1992). Ausgangspunkt des Entwurfs von Geräten der künstlichen Intelligenz müssen die theoretischen Vorstellungen über Art und Ablauf geistiger Tätigkeiten des Menschen sowie über deren Abhängigkeit von äußeren Einwirkungen sein. Ferner müssen die zu erwartenden Wirkungen auf Mensch und Umwelt und deren Bewertung bei der Planung berücksichtigt werden.

Wenn die Konstruktionsaufgabe lösbar ist, wird es von den erkennbaren Auswirkungen der Lösung auf das Individuum wie auf die Gesellschaft abhängen, wieweit die Realisierung vertretbar ist. Auch vor der elektronischen Datenverarbeitung gab es mechanische, informationsverarbeitende Maschinen. Informationsträger waren etwa Kurvenkörper, Kurbeltriebe oder auch Lochstreifen oder Lochkarten. Sie waren aber auf einem bestimmten, eng begrenzten Verwendungszweck festgelegt. Erst mit dem Computer entstanden vom späteren Verwendungszweck unabhängige, universelle daten- und informationsverarbeitende Maschinen. Sie werden auch zur Realisierung von künstlicher Intelligenz benötigt. Dementsprechend muß sich eine auf die technische Komponente der künstlichen Intelligenz gerichtete Betrachtung mit der Informationstechnik befassen. Dazu sind auch die Anwendungsbedingungen und die zu erwartenden Wechsel- und Folgewirkungen sowohl auf Individuen wie auf Gesellschaftssysteme – einschließlich der ihnen zur Verfügung stehenden sonstigen Technik – einzubeziehen.

Veränderungen der Industriegesellschaft durch Informationstechnik

Die Industriegesellschaft hatte im Laufe ihrer 200-jährigen Entwicklung mit den Erfindungen der Dampfmaschine, des Elektromotors, des Telefons und vielen anderen technischen Nutzungen wissenschaftlicher Erkenntnisse die Lebenswelt des Menschen zwar völlig verändert, doch blieb die Position des Menschen als allein fähig zur bewußten Veränderung nach von ihm selbst vorgegebenen Zielen unumstritten (BMFT 1991).

Die Veränderungen der Lebenswelt und die dadurch hervorgerufenen sozialen Umwälzungen hatten sich in der Mitte unseres Jahrhunderts auf einen gewissen Ausgleich hin entwickelt. Nun aber beginnt mit der Informationstechnik eine Zeit neuer Veränderungen. Diese neue Technik der Informations- und Wissensverarbeitung verändert die Lebensbedingungen. Sie verändert aber auch die menschliche Stellung in der phylogenetischen Entwicklung.

Die nachindustrielle Gesellschaft kann – Daniel Bell hat schon darauf hingewiesen – die Bedürfnisse nach materiellen Gütern im wesentlichen befriedigen. Ihre Mangelerscheinungen sind dagegen in den Informations-, Koordinierungs- und Zeitkosten zu suchen. Wenn kein Mangel an materiellen Gütern mehr besteht, werden Information und Kommunikation, was die relative Knappheit anlangt, an die Stelle der traditionellen, das Wachstum bestimmenden Faktoren treten (Breitenstein 1983). In der heraufziehenden nachindustriellen Wirtschaft tritt die Güterproduktion hinter die Dienstleistungen zurück. Die Verarbeitung von Informationen und theoretischem Wissen gewinnt einen steigenden Anteil (Spur 1990). Nicht nur die an Entscheidungen beteiligten Menschen haben einen wachsenden Informationsbedarf, auch in Maschinen werden in zunehmendem Umfang Informationen fest installiert oder sie werden mit Informationen laufend oder in Abständen versorgt. Inhalt und Umfang der in Maschinen installierten Informationen und damit deren Abläufe und Zusammenhänge mit dem Gesamtsystem sind für den dort Tätigen und erst recht für die Außenstehenden nicht mehr erkennbar und darum auch nicht mehr verständlich. Darum fehlt vielfach auch die Lebenserfahrung, die Zusammenhänge unserer technisch zivilisatorischen Welt zu durchschauen. Das Wissen über die Welt, über unsere Kultur und

über Gesetzlichkeiten wächst in einem nicht mehr zu übersehenden Maße. Es ist dem einzelnen Menschen nicht mehr möglich, sich all das Wissen, das ihn angeht, anzueignen, im eigenen Bewußtsein zu halten und zu verarbeiten (Lübbe 1991).

Informationstechnik – Herausforderung für Gesellschaft und Staat

Die Überwindung dieser Schwierigkeiten ist eine Herausforderung für die nachindustrielle Gesellschaft und den Staat. Zielvorgaben, Wegweisungen, Entwürfe für Handlungsalternativen sind gefordert, Kontrollinstrumente, Grenzen müssen festgelegt, aber auch erforderliche Freiräume gesichert werden. Alle sind aufgefordert, an dieser Aufgabe mitzuwirken – Parteien, Sozial- und Tarifpartner, Wissenschaft und Technik, Künstler und Theologen. Die Gesellschaft muß diese Aufgabe aus ihrem Selbstverständnis heraus und nach ihren Erkenntnissen und Erfahrungen angehen. Gesellschaft wird als ein „umfassendes Sozialsystem aller kommunikativ füreinander erreichbaren Handlungen" verstanden (Kägi 1984). Maturana und Varela bezeichnen als Kommunikation die aus der sozialen Kopplung entstehende Koordination des Verhaltens menschlicher Gruppen (Maturana, Varela 1987). Technik und Gesellschaft sind in einem wichtigen Bereich dekkungsgleich, nämlich insoweit der sinnvolle Bezug für Technik die Arbeit ist und diese in Form der Vereinigung von Personen, d.h. der Arbeitsteilung, Gesellschaft bereits schon darstellt. Die Begriffe Technik und Gesellschaft bezeichnen Zustände, statische Dinge, ordnungspolitische Gleichgewichte, die auch den Wandel einbeziehen (Bornschier 1980). Diese Übereinstimmung in einem wichtigen Bereich legt es nahe, auch strukturelle Verwandtschaften und Ähnlichkeiten in den Abläufen zu vermuten und mit den Methoden der Gesellschaftswissenschaften technologische Zusammenhänge zu analysieren.

Die Soziologie sieht ihre Aufgabe darin, die Gesetze des gesellschaftlichen Zusammenlebens zu erforschen und Möglichkeitenzu entwickeln, das soziale Zusammenleben besser zu organisieren (Fleckenstein 1965). Sie befaßt sich heute zunehmend mit systemtheoretischen Interpretationen der Gesellschaft, Erklärungsmodellen, die sowohl der Analyse von Organisationen dienen wie

auch Zugang zum Verständnis technischer Phänomene verschaffen können. Das ist gerade für die Einordnung und Bewertung der modernen Regelungs-, Steuerungs-, Informations- und Kommunikationstechniken wichtig, weil sie in wachsendem Maße nicht nur den beruflichen, sondern auch den privaten Alltag einschließlich unsers Freizeitverhaltens prägen.

Die modernen Systemtheorien gehen bei ihrem Systembegriff von der System-Umwelt-Differenz aus. Systeme definieren sich durch ihre Grenzen. Damit ist nicht gesagt, daß es keine grenzüberschreitenden Prozesse oder Abhängigkeiten gäbe, nicht einmal, daß diese von geringerer Bedeutung wären als die innerhalb der Systeme. Das, was sich ändert, sind die „Werte" der über die Grenze ausgetauschten „Güter". Handelt es sich z.B. um eine Information, so ist typischerweise ihre Bedeutung in bezug auf ihre Akzeptabilität, ihr Verständnis, ihre Weiterverwendung, ihren Neuheitswert und die von ihr ausgelösten Handlungs- oder Erwartungsfolgen „vor" und „hinter" der Grenze nicht dieselbe (Hahn 1987). Die innere Gliederung eines Systems kann auf zwei Weisen analysiert werden, einmal kann man es in seine Elemente zerlegen und Zahl und Art der Verknüpfungen feststellen. Die andere Form der theoretischen Zerlegung des Systems geht von dessen Funktionen aus. Als komplex werden solche Systeme bezeichnet, bei denen es nicht mehr möglich ist (z.B. auf Grund der Zahl), jedes Element mit jedem anderen zu verknüpfen. Eine Kombination ist dann nur eine von verschiedenen Möglichkeiten. Sie ist insofern also nicht zwingend, sondern kontingent und deshalb stets riskant.

Systemtheorie der Gesellschaft

Wie immer man die von verschiedenen Seiten heftig bestrittene Legitimität dieses Ansatzes einschätzt, er hat jedenfalls für die Soziologie der Gegenwart repräsentative Bedeutung. Vertreter systemtheoretischer Gesellschaftskonzeptionen sind Talcott Parsons und Niklas Luhmann. Parsons geht es darum, eine Theorie zu überwinden, die das Handeln aus der Annahme von fixen Bedürfnissen und situativen Möglichkeiten ableiten möchte. Für ihn tritt als erstes an die Stelle des artspezifischen konstanten Bedrüfnisses die sozio-kulturell erworbene Bedürf-

nisdisposition. Als zweiter Bereich von Selektionsregeln müssen moralische Normen angenommen werden. Auch hier wirken kulturell vermittelte Muster des Empfindens und des Urteilens, es gibt Werte, anhand derer auf kulturell variable Weise über die Frage entschieden wird, woran man Wahres von Falschem unterscheiden kann (Normkonsens).

Die elementare Einheit eines Sozialsystems ist für Parsons die einzelne Handlung. Der Modus der Integration der Einheiten zu einem System ist jeweils ein anderer. Persönlichkeits- und Sozialsystem sind nicht als Teile eines Ganzen verknüpft, sondern als Durchdringungsverhältnis von Systemen, die für einander wechselseitig Umwelten darstellen, zu sehen. Ein soziales System hat vier Hauptfunktionen zu erfüllen: 1. Umweltanpassung, 2. Zielerreichung, 3. Integration und 4. Legitimation. Jedes soziale System muß auf irgendeine Weise eine Lösung für diese Probleme finden, wenn es über längere Zeit in einer Umwelt „überleben" soll (Parsons 1976). Dabei ist aber zu bedenken, daß ein Subsystem seinerseits nur bestandsfähig ist, wenn es selbst auch die vier Grundfunktionen wahrzunehmen vermag. In modernen Gesellschaften übernimmt die Funktion der Umweltanpassung des Systems primär das Subsystem Wirtschaft, die Funktion der Zielbestimmung und der Zielerreichung vor allem das Subsystem Politik, die Funktion der Integration obliegt den Sozialinstanzen und die Legitimation dem Rechtssystem.

Parsons Ansatz kann zwar die funktionale Beziehung eines Subsystems zum Gesamtsystem erklären. Die Funktion des Gesamtsystems aber muß er als gegeben voraussetzen. Eine solche Erklärung ist mit dem Ansatz von Luhmann jedoch möglich. Für ihn ist jede Umwelt aus der Perspektive des Systems komplexer als es selbst. Es gibt nur eine beschränkte Zahl möglicher Zustände, die es annehmen kann. Ein System kann nicht auf jede Zustandsänderung der Umwelt mit einer eigenen antworten. Es muß also gegenüber sehr vielen Umweltänderungen indifferent bleiben. Die Funktion von Systemen gegenüber der Umwelt besteht insofern immer in Reduktionen von Komplexität. Für die Analyse von individuellen und sozialen Systemen wählt Luhmann „Sinn" als Grundbegriff. Sowohl Bewußtsein wie Kommunikationssysteme (= soziale Systeme) sind Sinnsysteme. Ihre Grenzen setzen sie selbst. Auch ihre Festlegung geht von

einem „Sinn" aus. Sozialsysteme liegen für Luhmann nicht erst bei Normkonsens vor, sondern bereits dann, „... wenn Handlungen mehrerer Personen sinnhaft aufeinander bezogen werden und dadurch in ihrem Zusammenhang abgrenzbar sind von einer nicht dazugehörigen Umwelt. Sobald überhaupt Kommunikation unter Menschen stattfindet, entstehen soziale Systeme" (Luhmann 1984). Hier wird somit deutlich: Kommunikation begründet Sozialsysteme. Das heißt für die Kommunikation und die dazu verwandten technischen Einrichtungen, daß sie der Eigenart und den Erfordernissen sozialer Systeme gerecht werden müssen. Eine qualitative Verbesserung der Kommunikation drückt sich in erweiterter Kompetenz und Koordinations- und Selbstorganisationsfähigkeit des sozialen Systems aus (Müller 1990). Solche Verbesserungen sollen auch „Künstliche Intelligenz" und „Expertensysteme" bringen.

Informationstechnik als soziotechnisches System

Die hohen Erwartungen an die Künstliche Intelligenz haben sich bisher nicht erfüllt. Der Ansatz, Systemtechnik und Wissen in Regeln zu fassen und nach einer klassischen Logik, etwa der Prädikatenlogik auszuwerten, hat noch nicht zum Erfolg geführt (Bonse 1991). Es kann daher nicht überraschen, daß man nach anderen Vorstellungen und Modellierungen von Intelligenz sucht. Es zeichnet sich ab, daß menschliche Entscheidungen nicht anhand von Regeln, sondern anhand von situativ vernetzten Wissenselementen getroffen werden (VDI 1991). Es regen sich zudem Zweifel, ob es tatsächlich in Einklang mit Wissenschaft und Erfahrung steht, wenn man sich Intelligenz und Wissen als lokale Eigenschaft eines Trägersubjekts vorstellt. Diese menschlichen Fähigkeiten haben sich im Verlauf der Evolution der Menschheit als soziales System herausgebildet (Phylogenese) und bilden sich auch individuell nur innerhalb sozialer Beziehungen heraus (Ontogenese) (Müller 1990). Die Intelligenz des Individuums allein reicht zur Sicherung der Existenz und zur Bewältigung der damit verbundenen Probleme nicht aus. Dazu bedürfte und bedarf es der kommunikativen Intelligenz in Organisationen, in sozialen Systemen. Die Informatik erkennt nunmehr auch die Grenzen ihrer vorwiegend naturwissenschaftlich-technisch fundierten Basis und erweitert ihr informationstechnisch orientiertes Systemverständnis in den sozio-

technischen Bereich. Das hat sowohl für die Informationstechnik und die von ihr zu gestaltenden technischen Systeme wie für die Gesellschaft und ihre Subsysteme Folgen.

Wenn die Maschinen dem soziotechnischen Systemverständnis gerecht werden sollen, müssen sie nach Bedarf relevante Merkmale aus ihrer Umwelt aufnehmen können. Das wird wahrscheinlich erfordern, daß die Maschine eine in irgendeinem Sinne vollständige Repräsentation mit ihrer Umwelt hat. Dabei genügt es nun sicher nicht, ein starres Abbild der Umwelt zu speichern, etwa nach Art der Kamera oder des Tonbandes. Vielmehr muß diese Repräsentation die Umwelt zergliedern, in sinnvolle Einheiten zerlegen, muß diese Einheiten nach allgemeinen Gesichtspunkten klassifizieren, muß vertraute Situationen und Muster wiedererkennen können und vieles mehr (Klotz 1988).

Die Maschine muß ein Symbolsystem besitzen, das fähig ist, Umwelt darzustellen. Die einzelnen Symbole müssen dabei mehr sein als reine strukturlose Zeichen, Stellvertreter für den dargestellten Gegenstand. Sie müssen diesen vielmehr in seiner Substanz, seiner Struktur, wiedergeben. Diese Symbole müssen die Basis für einen Organisationsprozeß abgeben, der einerseits Situationen der Umwelt modellartig abbildet und speichert, andererseits neue sinnvolle Situationen aufbaut (Malsburg 1987). Damit fällt solchen Maschinen eine Gestaltungsaufgabe zu, von der wir nicht von vornherein sicher sein können, daß die Ergebnisse in unserem Interesse liegen. Von der Existenz solcher Ergebnisse gehen aber Wirkungen aus. Es wäre also wichtig, die Folgen von solchen Systementwicklungen vorher abzuschätzen zu können.

Risikodialog, Technikfolgenabschätzung

Seit der Einführung der Wahrscheinlichkeitstheorie ist man in der Lage, verschiedenartige Handlungsalternativen als Verteilungen von Eintrittswahrscheinlichkeiten und zuzuordnenden Ergebnissen darzustellen und Entscheidungen zuzuordnen (Erdmann 1990). Wenn auf diese Weise auch die Auswahl einer Alternative auf eine rationale Grundlage gestellt werden kann, so ist die Realisie-

rung damit noch keineswegs gesichert. Für die Durchsetzbarkeit ist nicht allein das Ergebnis von Risikoanalysen maßgeblich. Vielmehr bedarf ein Handlungsprojekt der Akzeptanz der Betroffenen.

Verschiedenartige Interpretationen von scheinbar gleichwertigen Risiken hängen oft mit der Zugehörigkeit zu bestimmten gesellschaftlichen Gruppierungen und deren spezifischer Sicht der Welt zusammen. Daher sind die Verhaltens- und Verfahrensweisen von gesellschaftlichen Subsystemen, wie Wirtschaft, Politik, Kirchen, Frauenbewegung usw., zu beachten. Die Bedingungen für einen „Konsens" müssen geschaffen werden. Dabei sollte sich die Auseinandersetzung von der Illusion freimachen, das jeweilige Risiko beherrschen zu können, und sich darum bemühen, es durch ein professionelles Management auf der Grundlage eines Risikodialogs der gesellschaftlichen Gruppen und Interessen zu bewältigen. Eines solchen Dialogs bedarf auch die Technikfolgenabschätzung (Haller 1990). Sie ist aber bei der Informationstechnik besonders schwierig, weil diese prinzipiell unvollständig ist und zur Vervollständigung des Menschen und der Berücksichtigung des Anwendungsfalles bedarf. Unter diesen Umständen ist Folgenabschätzung nicht sonderlich erfolgversprechend. Sinnvoller dagegen erscheint eine auf demokratisch verantworteter Planung beruhende Technikbewertung und Technikgestaltung (Steinmüller 1987).

Rammert schlägt vor, die Analyse der Technisierung von Kommunikation wie eine Medienanalyse zu betreiben. Damit rücken Fragen des kulturellen Wandels in den Blickpunkt und weisen auf die Funktionen als Wissenspeicher, als kombiniertes Wissens- und Kommunikationsmedium sowie als Medium zur Organisierung sozialer Handlungen hin. Nicht mehr das fachliche oder professionelle Wissen der einzelnen, sondern der Zugang zum neuen Medium würde die Machtposition der zu einer Organisation Gehörigen bestimmen (Rammert 1990).

Dieser Vorschlag erscheint umso berechtigter, weil in den nächsten Jahrzehnten die Computertechnik, die klassichen Druck-Medien sowie die Bereiche des Radios, Fernsehens und Videos zusammenwachsen werden – „sinn-bildlich" dargestellt in der einheitlichen Kompaktdisk für Computerdaten, Musik, Bilder und Filme. Insgesamt dürfte die moderne Informationstechnik als „kulturtechnisches

Werkzeug" ebenso epochale Veränderungen einleiten wie die Entwicklung der Schrift und die Entwicklung der Buchdruckerkunst es getan haben (BT 1989).

Technikbewertung

In dem zur Technikbewertung erforderlichen demokratischen Dialog tragen die Ingenieure und ihre Vertretung in der Gesellschaft, der Verein Deutscher Ingenieure, eine besondere Verantwortung. Ingenieure können durch Entwurf und Realisierung das technisch Mögliche zeigen. Ihr Fachwissen befähigt sie, erreichbare Ziele in das Bewußtsein der Öffentlichkeit zu rücken. In der Auseinandersetzung über neue Technologien müssen alle gesellschaftlichen Bereiche zusammenarbeiten und die aus ihrer Sicht wichtigen Aspekte einbringen. Ethische und moralische Leitlinien sollten von allen an der Entwicklung Beteiligten anerkannt werden. Das erfordert einen Konsens über ein verbindliches gesamtgesellschaftliches Zielsystem, das die Möglichkeiten der Technik und auch ihre gesellschaftliche Weiterentwicklung im Auge behält.

Aus dieser Sicht hat der VDI eine Richtlinie – VDI 3780, Empfehlungen der Technikbewertung – erarbeitet. Sie soll dazu helfen, Bewertungsprozesse vor einem breiten Bewertungshorizont im gesellschaftlichen Dialog zu führen und möglichst alle Folgen einer Technik für Umwelt und Gesellschaft nach außertechnischen und außerwirtschaftlichen Werten zu beurteilen. Der Bewertungsprozeß bleibt so nicht auf einen einzelnen Entscheidungsträger beschränkt, sondern wird von einem Netzwerk gesellschaftlicher Einrichtungen vorbereitet, unterstützt und begleitet (VDI 1991b).

Gesetzgebung, Rechtsprechung, staatliche Kontrolle

Der gesellschaftliche Dialog muß am Ende Entscheidungsgründe für die Politik liefern. Ihre Aufgabe ist es, die für das Gemeinwesen, den Staat, das Land, eine Gemeinde, u.U. auch für internationale Gemeinschaften alle für deren Bestand entscheidenden Sach- und Bewußtseinsbereiche zu gestalten. Politisches Handeln vollzieht sich in der Regel innerhalb gewisser durch Konvention oder Sat-

zung festgelegter Normen wie Völkerrecht, Verfassung, Gesetze, Statuten usw. (Gablenz 1965).

Neue Technologien und erst recht solche, die die Position des Menschen in seiner Umgebung und in der Technik verändern können, stoßen auf Mißtrauen oder gar auf Ablehnung. Wenn zudem die Folgen solcher Techniken nicht hinreichend zutreffend beurteilt werden können, so wird die heute unverzichtbare Akzeptanz nur erreicht werden können, wenn sichergestellt ist, daß Grenzen nicht voreilig überschritten werden, bevor nicht alle verfügbaren Sicherungen für den Gefahrenfall getroffen sind. Um das zu gewährleisten und dafür Sorge zu tragen, daß Sicherheitsmaßnahmen, so lange sie erforderlich sind, aufrecht erhalten werden, müssen Gesetze für ihre Anbringung und zur Kontrolle ihrer Wirksamkeit erlassen werden. Gesetze sind für die Akzeptanz neuer Technologien eine wichtige Voraussetzung.

Mit der Anwendung der Informationstechnik können Maßnahmen erforderlich werden, die die Freiheitsrechte der im jeweiligen Bereich Tätigen berühren können. Wenn technische Systeme zu Nervenzentren der Gesellschaft werden, muß deren Betrieb unter allen Umständen gewährleistet werden. In einem solchen Konflikt können durch die Notwendigkeit betriebsinterner Sicherungssysteme eine Reihe von Grundrechten Wandlungen erfahren. Als Beispiele sind die Gefahren für den Schutz der Menschenwürde, des Grundrechts auf Privatsphäre, der informationellen Selbstbestimmung, der Handlungsfreiheit, des Persönlichkeitsrechts und schließlich des Streikrechts zu nennen (Rumpf 1972). Solche Gesichtspunkte bedürfen bei der Gesetzgebung rechtlicher Würdigung. Bei der Konzeption des Informationsrechts über die Verfügbarkeit, Richtigkeit und Ungefährlichkeit von Information muß von vornherein bewußt sein, daß es in ihm nicht um die Information als solche, sondern um die Bedeutung der Information für den Menschen in der Gesellschaft geht, nicht die „Informationsgesellschaft", sondern die – richtig – „informierte Gesellschaft" steht im Vordergrund. Bezeichnungen wie „Datenschutz", „Datenverkehrsordnung" oder „Informationsverfassung" betonen zwar zu Recht die Bedeutung von Daten und Information, vernachlässigen jedoch leicht den instrumentalen Charakter des Informationsrechts – wie des Rechts überhaupt – als Dienst am Menschen und an

der Gerechtigkeit (Sieber 1989). Es bedarf auch hier der Einbeziehung eines soziologischen Ansatzes, der die Veränderung gesellschaftlicher Steuerungsfunktionen, -leistungen und -defizite behandelt und in dem die Funktionen von Recht unter dem Aspekt von „Verrechtlichung und Entrechtlichung" diskutiert werden (Lennartz 1989). In der Rechtsprechung muß der Staat die gesellschaftlichen Funktionen des Rechts zur Geltung bringen. Die Aufgabe des Staates endet aber nicht bei der Gesetzgebung und Rechtsprechung. Der dritten Staatsgewalt, der Exekutive, obliegt es, für eine uneingeschränkte, sachgerechte und sozialverträgliche Anwendung der Gesetze zu sorgen. Das gilt in einem ganz spezifischen Sinne für die Durchführung von Gesetzen für den technischen Bereich. Dem Staat als Freiheits- und Organisationsgaranten obliegt die Kontrollaufgabe, der Bürger als Freiheitsberechtigter und Rechtsverpflichteter muß die Kontrolle dulden. Sie muß nicht immer vom Staat selbst oder unmittelbar ausgeübt werden. Maßgebender und letztlich entscheidender Kontrolleur ist aber der Staat. Private Stellen können mit Selbstkontrolle die Aufgabe des Staates erleichtern, aber nicht ersetzen.

Verantwortung von Entscheidungsträgern und Handelnden

Gesetze und politischer Dialog können letztlich aber nur wirksam werden, wenn alle in Technik und Politik Entscheidenden und Handelnden sich der Folgen ihres Handelns für Gesellschaft und Umwelt bewußt sind. Die Fähigkeit zur Verantwortung – zu ethischem Verhalten – beruht in der Befähigung des Menschen, zwischen Alternativen des Handelns mit Wissen und Wollen zu wählen. Verantwortung ist also komplementär zur Freiheit. Sie ist die Bürde der Freiheit eines Tatsubjekts: Wir sind verantwortlich mit unserer Tat als solcher (ebenso wie mit ihrer Unterlassung), und das gleichviel, ob jemand da ist, der uns – jetzt oder später – zur Verantwortung zieht (Jonas 1987). Diese Verantwortung obliegt dem einzelnen wie auch der Gesellschaft. Sie sollte ethische Werthaltungen auch institutionell absichern. Dabei soll zwar der Staat den Rahmen setzen; die Entscheidungsträger, die Handelnden und die von dem Ergebnis geplanten Handelns Betroffenen sollten aber selbst die möglichen Folgen abschätzen aus dem Bewußtsein, daß die umweltverträglichen und gesellschaftlich gewünschten

Ergebnisse auch die langfristig wirtschaftlich und sozial erfolgreichen sein werden (Hastedt 1991).

Literatur

BMFT (1991). Neurobiologie/Hirnforschung – Neuroinformatik, Künstliche Intelligenz. Bonn: Der Bundesminister für Forschung und Technologie (April 1991).

Bonse, E. (1991). Künstliche Intelligenz besinnt sich auf Datenverarbeitung zurück. VDI Nachrichten Nr. 39 (27.9.1991).

Bornschier, V. (1980). Technik und Gesellschaft. In: Technik wozu und wohin? Zürcher Hochschulforum, Bd. 3. Zürich und München: Artemis Verlag.

Breitenstein, H. (1983). Auswirkungen der Informationsverarbeitung auf die Gesellschaft von morgen. IBM Nachrichten 33, Heft 267.

BT (1989). Bericht der Bundesrepublik über die Auswirkungen der Urheberrechtsnovelle 1985 und Fragen des Urheber- und Leistungsschutzrechts. BT Drucksache 11/4929 vom 7.7.89, S. 23 ff.

Erdmann, G. (1990). Rationale Risikoaversion. Neue Zürcher Zeitung Nr. 24 (31.1.1990).

Fleckenstein, J.O. (1965). Naturwissenschaft und Technik. München: Verlag Georg D. W. Callwey.

Gablenz, H.O. von der (1965). Einführung in die Politische Wissenschaft. Köln und Opladen: Westdeutscher Verlag.

Hahn, A. (1987). Die Gesellschaft als System. Neue Zürcher Zeitung Nr. 193, (23./24.8.1987).

Haller, M. (1990). Der „Risikodialog" als Chance. Neue Zürcher Zeitung Nr. 24 (31.1.1990).

Hastedt, H. (1991). Aufklärung und Technik. Frankfurt: Suhrkamp Verlag.

Jonas, H. (1987). Vom Maßhalten im Gebrauch der Macht zum Maßhalten im Erwerb der Macht. Handelsblatt Nr. 194 (9./10.10.1987).

Kägi, E.A.(1984). Homo functionalis. Neue Zürcher Zeitung (8.2.1984).

Klotz, K. (1988). Die Illusion von der Künstlichen Intelligenz. Neue Zürcher Zeitung Nr. 86 (27.4.1988).

Krämer, S. (1992). Eine weitere kopernikanische Wende? (Dieser Band)

Lennartz, H.-A. (1989). Probleme der Techniksteuerung durch Recht – am Beispiel des bun?desdeutschen Datenschutzrechts. Recht der Datenverarbeitung 1989, Heft 5/6.

Lübbe, H. (1991). Zur moralischen Verfassung der wissenschaftlich-technischen Zivilisation. In: Fraunhofer-Institut für Materialfluß und Logistik 1981 - 1991, 10 Jahre IML.

Luhmann, N. (1984). Soziale Systeme. Frankfurt: Suhrkamp Verlag.

Malsburg, C. von der (1987). Kann der Mensch intelligente Maschinen bauen? Stahl und Eisen (1987) Nr. 24.

Maturana, H.R., Varela, F.J. (1987). Der Baum der Erkenntnis. Bern-München-Wien: Scherz Verlag.

Müller, R.A. (1990). Selbstorganisation und verteilte Intelligenz. Vortrag beim Symposium „Evolutionäre Wege in die Zukunft" des Instituts für Zukunftsstudien. Berlin, 29. - 30.11.1990.

Parsons, X., Talcott, Y. (1976). Zur Theorie sozialer Systeme. Hrsg. S. Jensen, Opladen: Westdeutscher Verlag.

Rammert, W. (1990). Technikgenese und die Einsatz von Expertensystemen aus sozial?-wissen?schaftlicher Sicht. Vortrag zur Auftaktveranstaltung des Verbundprojekts „Veränderung der Wissensproduktion und Wissensverteilung durch Expertensysteme" am 2. Mai 1990 im VDI-Zentrum Düsseldorf.

Rumpf, H. (1972). Technik und Bildung. VDI-Zeitschrift 114 (1972) Nr. 17, Dezember.

Sieber, U. (1989). Informationsrecht und Recht der Informationstechnik. Neue Juristische Wochenschrift, Heft 41, 42. Jahrgang, 11. Oktober 1989.

Spur, G., Steusloff, H. (1990). Information als Produktionsfaktor. Fraunhofer Gesellschaft, Jahresbericht 1990. München: Fraunhofer Gesellschaft.

Steinmüller, W. (1987). Technologiefolgenabschätzung. Computermagazin 12/87.

VDI (1991a). VDI Nachrichten: Künstliche Intelligenz lernt hinzu. VDI Nachrichten Nr. 39 (27.9.1991).

VDI (1991b). VDI-Richtlinien: VDI 3780 – Empfehlungen zur Technikbewertung.

Mögliche Auswirkungen einer entwickelten KI auf Arbeits- und Lebenswelt

G. Görz, A. Kremeier, H. Röpke, P. Schreiber, G. Strube, I. Wachsmuth, M. Wilker

Hypothetische Stadien einer entwickelten KI

In diesem Beitrag wird der Versuch unternommen, ausgehend von einer Extrapolation gegenwärtig realistischer KI-Systeme, hypothetische – zunehmend fiktive – Stadien zukünftiger KI-Entwicklungen zu skizzieren und im Hinblick auf ihre Implikationen für Arbeits- und Lebenswelt zu diskutieren. Die Charakterisierung von Stadien einer entwickelten KI setzt eine definitorische Abgrenzung von herkömmlichen komplexen Softwaresystemen voraus. Als Mindestvoraussetzung für eine entwickelte KI soll hier der weitverbreitete Einsatz von Systemen gelten, die Produkte – und keine Entwicklungsprototypen – sind. Der Begriff „Künstliche Intelligenz" soll im weiteren beinhalten, daß derartige Systeme

- über kognitive Komponenten verfügen (optische/akustische Wahrnehmung, etc.)
- logische und bereichsspezifische Schlußfolgerungen ziehen können
- heuristische Problemlösungsstrategien entwickeln
- diese Komponenten integrieren (holistischer Ansatz)

Unter Berücksichtigung dieser Definition ist zumindest fraglich, ob die Bezeichnung „KI" für gegenwärtige Systeme angemessen ist. Jedoch erscheint die Realisation solcher Systeme zunehmend wahrscheinlich.

Stadium 1: denkbar etwa in den nächsten 10 Jahren

Maschinen formalisieren einen abgrenzbaren und in sich einigermaßen homogenen Teil der Wirklichkeit: Sie übernehmen intellektuelle Routinetätigkeiten wie Rechnen, Ordnen und Zuordnen, Korrelation digitaler Zeichenfolgen, Entscheidungsunterstützung. Hierzu gehören einerseits „klassische" Expertensysteme wie sie etwa bereits heute in der Diagnostik technischer Systeme ansatzweise eingesetzt werden, andererseits Systeme, die „intelligente" Interpretationen von Meßdaten (wie etwa UV-, NMR- oder EEG-Spektrenklassifizierungen) leisten, oder Systeme, die nach vorgegebenen Ablaufrichtlinien und biometrischen Rahmenbedingungen Versuchsplanungen vornehmen.

Stadium 2: denkbar vielleicht in den nächsten 30 Jahren

Maschinen übernehmen in größerem Umfang selbständig Wissensrepräsentation und Wissensverknüpfung. Sie können auch aus unscharfem Wissen zu verwertbaren Ergebnissen kommen, können Fehler diagnostizieren, selbst Reparaturen ausführen oder den Service herbeirufen. Möglich scheinen Durchbrüche für automatische Sprachübersetzungssysteme sowie entwickelte Unterstützungssysteme für die Entwicklung von DV-Programmen. Starkes Vordringen von KI-Systemen in den Freizeit- und Bildungsbereich und in die allgemeine Informationstechnik.

Stadium 3: über 30 Jahre hinaus extrapoliert

Maschinen sind nur noch teilweise determiniert. Sie können Situationen, mit denen sie konfrontiert sind, absichtsvoll verbessern und umgestalten, indem sie aus sich selbst als Reaktion auf die Einwirkung der Umwelt intelligentes Verhalten zeigen. Wenn die Erforschung chaotischer Systeme und der Strukturbildung zu verwertbaren Ergebnissen führt, könnten solche Maschinen Spontaneität, Zielstrebigkeit, Kreativität und Sensibilität besitzen. Szenarien, die sich hierzu in der gegenwärtigen Literatur finden, betreffen autonome mobile selbstlernende

Systeme, „künstliches Leben" als Paradigma und als kühne Perspektive Moravecs „genetic takeover"[1].

Im folgenden betrachten wir die mit diesen hypothetischen Stadien einer entwikkelten KI verbundenen Problemfelder weniger in Erwartung ihres wahrscheinlichen Eintretens, als vielmehr hinsichtlich der Implikationen und der Wünschbarkeit solcher Zukunftsszenarien. Auch wenn die Möglichkeit der im dritten Stadium angesprochenen Entwicklungen bereits jetzt mit Skepsis betrachtet wird, darf nicht übersehen werden, daß die mit derartigen Annahmen verbundenen Zukunftsvisionen durchaus Einfluß auf Entscheidungsträger, etwa hinsichtlich der Finanzierung von Forschungs- und Entwicklungsprogrammen nehmen können.

Entsprechend dem Titel dieses Abschnitts untersuchen wir einzelne Fragenkomplexe einmal fokussiert auf die Veränderung von Qualifikationen und Arbeitsstrukturen und zum anderen hinsichtlich denkbarer Entwicklungen in der Lebenswelt des Alltags. Einzelbeiträge von Autoren des vorliegenden Abschnitts werden sich gesondert mit der Verantwortung von KI-Produkten und ihren potentiellen Gefahren befassen.

Veränderung von Qualifikationen und Arbeitsstrukturen

Bereits mit den für das erste Stadium skizzierten Entwicklungen können Probleme der Entwertung von Qualifikationen verbunden sein wie auch eine stärkere Schematisierung und Bürokratisierung beruflicher Abläufe. Auch wenn es sich dabei grundsätzlich um die gleichen Probleme wie bei entwickelter Informationsverarbeitungstechnik handelt, so könnte sich die Realisierungsmöglichkeit

[1] Moravec liefert eine Prognose der Zunahme an Rechenkapazität und prognostiziert die „human equivalence" von Maschinen in ca. 40 Jahren: "...sooner or later our machines become knowledgeable enough to handle their own maintenance, reproduction and self-improvement without help. When this happens, the new genetic takeover will be complete. Our culture will then be able to evolve independently of human biology and its limitations (...). A mind would require many modifications to operate effectively after being rescued from the limitations of the mortal body (...)." (Moravec 1988, S.3ff)

der skizzierten Systeme erst durch Einbezug von KI-Techniken ergeben. Die Wissensübertragung von einem Automaten auf den anderen ist denkbar einfach im Vergleich zu mühsamen Lern- und Wiederlernvorgängen, denen sich Menschen unterziehen müssen. *Wird eine umfangreiche Neubewertung gegenwärtiger Qualifikationen eintreten? Welche Arbeit bleibt dem Menschen vorbehalten, und wie wird sie bewertet? Werden sich grundsätzlich neue Strukturen in der Arbeitswelt entwickeln?*

Mit der in der Expertensystemtechnologie angestrebten maschinellen Speicherung und Vervielfältigung des Fachwissens qualifizierter Arbeitskräfte zeichnet sich eine Neubewertung der Ressource „Wissen" ab. Ökonomisch betrachtet ist Wissen ein knappes, aber sehr produktives Gut. Eine hohe Qualifikation hat einen hohen Preis. Maschinell erzeugte Verfügbarkeit von Wissen und damit Qualifikation wird den Preis für Durchschnittsqualifikationen sinken lassen. Dieser Wertverfall von Qualifikationen irritiert das Selbstverständnis derer, die sie mühsam erworben haben.

Der Befürchtung einer solchen Aushöhlung des Fachwissens durch Expertensysteme steht die Klage über ein sich ausbreitendes Spezialistentum gegenüber. In zunehmendem Maße führt der vom Arbeitsmarkt ausgehende Selektionsdruck dazu, daß in der beruflichen Qualifizierung die Aneignung sehr speziellen Wissens übermäßiges Gewicht erhält. Dies führt zur vielbeklagten Kommunikationsunfähigkeit der Experten verschiedener Fachgebiete einer Disziplin untereinander. Hier liegt möglicherweise eine Chance für Expertensysteme, dadurch Abhilfe zu schaffen, daß breit gebildete Fachkräfte in Spezialfragen unterstützt werden. Expertensysteme stellen komprimiertes Wissen dar. Sie „nehmen das Wissen nicht weg", sondern stellen es in aufbereiteter Form zur Verfügung. Der Einsatz von KI-Systemen in der Praxis – etwa in der technischen Diagnostik – zeigt, daß durch KI bislang Experten nicht „wegrationalisiert" werden.

Für die ersten beiden Stadien ist sicherlich anzunehmen, daß eine sinnvolle Kooperation von Experten und Expertensystemen zustande kommt, die allerdings wesentlich von der Entwicklung intelligenter Mensch-Maschine-Schnittstellen

abhängen wird. Mittelfristig ist ein neuer Typus des Wissenschaftlers oder Ingenieurs denkbar, der weniger als gegenwärtig Spezialist zu sein genötigt ist. Darüber hinaus scheint die Praxis zu zeigen, daß sich die Benutzer beim Umgang mit derartigen Systemen schnell das darin aufbereitete Wissen aneignen.

In der Arbeitswelt wird KI vor allem deshalb eingesetzt, weil man sich Rationalisierungs-erfolge (z.B. durch Qualitätsverbesserung) verspricht. *Wird ein massiver KI-Einsatz im industriellen Sektor Arbeitslosigkeit erzeugen?* Hierzu schreibt der KI-Wissenschaftler Jörg Siekmann: „Die informationsverarbeitende Technologie – und deren schillerndstes Kind, die Künstliche Intelligenz – vernichtet Arbeitsplätze, und dieser Prozeß wird sich in den nächsten Jahren noch erheblich beschleunigen. Durch diesen Prozeß werden Millionen von Arbeitern und Verwaltungsangestellten zunächst das verlieren, was ihren ‚Marktwert' und nicht zuletzt ihr Selbstverständnis ausmacht, nämlich ihre Qualifikation, die nun nicht mehr gebraucht wird, und sie werden schließlich im großen Heer der ‚nicht mehr vermittelbaren' Arbeitslosen landen" (Siekmann 1989, S. 128). Den Versuch, eine Arbeitslosigkeit produzierende Informationstechnologie zu stoppen, indem man die Suche nach technologischen Innovationen unterbindet, hält Siekmann für illusorisch und auch unter ethischen Gesichtspunkten für fatal. Er weist darauf hin, daß beim gegenwärtigen Stand der Produktionstechnik noch viele unmenschliche Arbeiten anfallen: „Der Roboter, der einen Arbeiter ersetzt, führt eine Arbeit aus, die an Stumpfsinn und Brutalität an die der römischen Galeerensklaven erinnert und die von niemandem freiwillig ausgeübt würde" (Siekmann 1989, S. 128). Das von Siekmann prognostizierte Heer der Arbeitslosen könnte sich jedoch im günstigen Falle ebenso als Übergangsphänomen erweisen wie im Gefolge der ersten Industriellen Revolution. Mittelfristige Auswirkung könnte weitere Produktivitätssteigerung und damit einhergehende Reduktion individueller Arbeitszeit sein.

Auf lange Sicht, sicherlich für das hypothetische dritte Stadium, wäre zu erwarten, daß der Arbeitsprozeß zumindest in der Güterproduktion in einem Maße automatisiert wird, das dem Menschen ein nahezu vollständiges Heraustreten aus diesem Prozeß gestattet. Dabei besteht allerdings die Gefahr, daß Ethik und Verantwortung zu bloßer Haftbarkeit degenerieren. Arbeit wird immer weniger ein

Hantieren mit physischen Objekten beinhalten. Die künftigen Qualifikationsanforderungen könnten sich in zunehmendem Maße das Wissen um Kontexte, in denen maschinelles Wissen relevant ist, erstrecken. Damit ginge ein Wandel des Begriffs von Arbeit einher, der sich im Extremfall auf bloßes Entscheiden und vor allen Dingen – da Entscheiden wiederum maschinenunterstützt stattfinden wird – Verantworten verlagern könnte. Die räumlich und zeitlich potentiell unbegrenzte Verfügbarkeit von Information, gekoppelt mit dem Entstehen neuer Kommunikationsmedien, reduziert die Distanz zwischen Arbeit und dem Bereich des Sozialen. Arbeit könnte sich dann in einer Art „Zwitterform" zwischen Produktion und Interaktion darstellen.

Werden durch solche Folgen des KI-Einsatzes grundlegende Sinnfragen der Menschheit berührt? Individuen erfahren Sinn wesentlich dadurch, daß sie schöpferisch arbeiten können. Wenn Maschinen, und das wäre für das dritte Stadium nicht auszuschließen, Rationalität, Kreativität und Spontaneität künstlich erzeugen könnten, würden herkömmliche Formen der Arbeit und deren positive Auswirkungen auf die persönliche Lebensführung teilweise eliminiert. Dies würde mindestens zu einer grundlegenden Neudefinition von Arbeit, wahrscheinlich aber zu beträchtlichen sozialen Veränderungen führen.

Fragen, auf die sich bisher keine Antworten geben lassen, ergeben sich für die Einschätzbarkeit und Verantwortung der durch den Einsatz von KI-Technologie denkbaren Rollenumverteilungen in der Arbeit. *Welche Risiken ergeben sich aus der mangelnden Durchschaubarkeit der Maschine und ihres Arbeitsergebnisses? Welche Gefahren können aus der fehlenden Verantwortung einer Maschine für ihre Arbeitsergebnisse folgen?* Wenn etwa Reparaturen, wie für das zweite Entwicklungsstadium postuliert, automatisch veranlaßt werden können, wer würde bei fehlerhaft durchgeführten Reparaturen die Verantwortung für die Folgen tragen? Ebenfalls problematisch erscheint die Abhängigkeit, in die sich der Mensch bei den angesprochenen Entwicklungen begibt. *Wie kann das Problem der sich verschärfenden Abhängigkeit von Maschinen bei der Entscheidungsfindung bewältigt werden?*

Zwar steht für „künstlich intelligente" Systeme fest, daß sie von Menschen geschaffen sind. Jedoch greift der Gedanke, daß alle ihre Handlungsmöglichkeiten von Menschen explizit vorgegeben und deshalb in ihren Implikationen abschätzbar sind, zu kurz: Technische Systeme sind – ebenso wie Menschen – fehleranfällig. Sie sind aber nicht geleitet von korrigierenden Haltungen und menschlicher Perspektive. Die komplexen Interaktionen der maschinellen Prozesse – *die auf die eigene Programmstruktur Einfluß nehmen können!* – erscheinen ebensowenig abschätzbar wie der Verlauf evolutionärer Prozesse. Der wichtigste Mangel derzeitiger – oder demnächst möglicherweise realisierter – Systeme scheint darin zu liegen, daß ihnen die Verankerung in einem körperlichen Bezugssystem und in erfahrungsgebildeten sozialen Kontexten fehlt, die als Voraussetzung für menschliches Urteilsvermögen und Verantwortung unverzichtbar erscheinen.

Denkbare Entwicklungen in der Lebenswelt

Im Alltag wird KI ebensowenig als „reine KI" auftreten wie menschlicher Geist als „reiner Geist" in Erscheinung tritt; es werden allenfalls Produkte mit „KI-Komponenten" entstehen oder solche, die auf der Basis von KI-Techniken realisiert sind. *Wie wird sich KI in das gesellschaftliche Leben integrieren? Ist überhaupt damit zu rechnen, daß KI-Produkte im Alltag besonders wahrgenommen werden, oder verlieren sie ihre „Identität"? Ist zu erwarten, daß aus der Verwertung von KI eine Generation von Konsumgütern entsteht, die die Lebenswelt drastisch verändert? Werden sie möglicherweise dazu führen, daß „künstliche Welten" neben einer realen Lebenswelt entstehen?*

In der gegenwärtigen Diskussion wird der KI-Technologie eine bedeutende Rolle in der modernen Informationsgesellschaft zugeschrieben und mit Vorhersagen über umbruchartige Entwicklungen der Alltagswelt in Verbindung gebracht, so etwa durch den Kommunikations-philosophen Vilem Flusser. Flusser (1987) geht davon aus, daß eines der wesentlichen Kennzeichen der Informationsgesellschaft die Redundanz von Information ist. Jeder kann jederzeit alles wissen und mit jedem über dieses Wissen kommunizieren. Durch die Möglichkeiten der KI

läßt sich Wissen maschinell neu kombinieren und kann so zu völlig neuen Ergebnissen führen. Gesellschaftliches Handeln wird als Resultat gesellschaftlicher Informationsverarbeitung durch Kommunikation dramatisch beschleunigt. Folgt man Flusser, steht die Menschheit vor einer Revolution unbekannten Ausmaßes: „... das ist mindestens so radikal anders, als der historisch lebende Mensch anders auf der Welt ist im Vergleich zum magisch-mythisch lebenden Menschen. Wir sind in einer Wende, die radikaler ist als die Achsenzeit Jaspers" (Flusser 1987, S. 123).

Ähnliche Überlegungen finden sich bei vielen KI-Enthusiasten und sind zugleich Motivation vieler Skeptiker und Warner, die befürchten, daß die „KI-Revolution" das Gesellschaftssystem „heißlaufen" läßt. Allerdings könnte hier die (oft latente) Annahme, Gesellschaften seien statische Systeme, leicht zu einer Überbewertung der KI-Folgen führen. Pointiert ausgedrückt ist KI kein Fluidum, das über eine mehr oder weniger wehrlose Sozialgemeinschaft „ausgegossen" wird, sondern sie findet auf dem Umweg über die (oft mühsame und langwierige) Produktentwicklung Eingang in gesellschaftliche Wirkungszusammenhänge. Soziale Systeme haben seit der industriellen Revolution weitgehende Aufnahmefähigkeit für technologischen Fortschritt bewiesen. Mittlerweile ist es gesellschaftlicher Konsens, daß eine sich rasch fortentwickelnde Technologie als selbstverständlich vorausgesetzt wird. In vielen Fällen werden KI-Produkte als verbesserte Substitute bereits in Gebrauch befindlicher Artefakte in Erscheinung treten. Dies ließe eine eher unspektakuläre soziale Integration von Auswirkungen einer entwickelten KI erwarten.

Um es prononciert zu formulieren: Es wird sicher mehr Überraschung hervorrufen, wenn in zehn Jahren keine zu 99 Prozent sprachverstehenden PC's mit erschwinglichen Preisen auf dem Markt sein werden, als wenn sie es wären. Vergleiche mit anderen technischen Innovationen mögen das verdeutlichen: Seit ca. fünf Jahren verdrängen Camcorder auf breiter Ebene herkömmliche Super-8-Kameras. Zehn Jahre zuvor ließen Taschenrechner die letzten Rechenschieber verschwinden – ein gewaltiger Sprung für den Ingenieur, aber für einen Achtklässler eher eine Bagatelle. Auch die Camcorder etablierten sich eher still und folgenlos. Trotz deren leichter Handhabbarkeit und eines relativ geringen Preises

ist die Gesellschaft weit davon entfernt, sich zu einem ständig filmenden Kollektiv von Autisten zu wandeln. Ähnlich wie in Produkte integrierte Mikroprozessoren praktisch nicht in Erscheinung treten, wird eine sich in Produkten manifestierende KI öffentlich kaum als solche wahrgenommen werden, sondern die Wahrnehmung der Produkte wird sich eher auf deren Gebrauchswert beschränken. Die produktimmanente technische Revolution wird als selbstverständlich vorausgesetzt und gerät aus dem Blickfeld.

Drastischer scheinen die möglichen Entwicklungsverläufe durch neuartige Medien, die das Potential haben, das kommunikative Geschehen sozialer Systeme umzugestalten und zu verändern (vgl. oben Flusser). Bereits jetzt entsteht hier unter Verwendung von KI-Komponenten eine neue Generation von „virtuelle Maschinen" genannten Konsumgütern, die erhebliche gesellschaftliche Auswirkungen (Freizeitverhalten, Individualisierung bis hin zum Verlust angemessener sozialer Interaktionsfähigkeit) haben könnten; vgl. dazu Jean Baudrillard (1989)[2]. Diese Auswirkungen werden weniger durch die ursprüngliche Intention der KI, menschliche Verstandesleistungen zu simulieren und technisieren, verursacht, sondern sie sind eine denkbare Folge der durch KI-Techniken in Verbindung mit fortgeschrittener Informations-technologie ermöglichten Simulationen künstlicher „Realitäten".

In jüngster Zeit ist unter Bezeichnungen wie „virtuelle Realität" oder „Cyberspace" eine Technik vorgestellt worden, durch die vom Computer Kunstwelten

[2] „Nicht umsonst nennen wir sie virtuelle Maschinen: denn sie halten das Denken auf immer in der Schwebe, im hypothetischen Anspruch auf ein totales Wissen. Der Akt des Denkens ist dabei ins Endlose hinausgezögert. Die künftigen Generationen werden die Frage nach dem Denken so wenig wie die nach der Freiheit stellen: sie werden gleichsam an ihren Sitzen festgeschnallt, das Leben wie in einem Luftraum durchqueren. Ebenso werden die Menschen der künstlichen Intelligenz, mit ihrem Computer verbunden, ihren Gedanken-Raum durchmessen. Der virtuelle Mensch, reglos vor seinem Computer, macht Liebe via Bildschirm und seine Vorlesungen per Telekonferenz. Er wird zum motorisch und wohl auch zerebral Behinderten – der Preis, den er zahlen muß, um operational zu werden. Wie Brillen oder Kontaktlinsen eines Tages zu integrierten Prothesen einer Gattung werden, die den Blick verloren hat, so wird einst – kann man befürchten – die künstliche Intelligenz samt technischem Zubehör die Prothese einer Gattung werden, der das Denken abhanden gekommen ist." (zitiert nach Waffeneder 1991, S. 287)

erzeugt werden, in die Benutzer mittels hochauflösender graphischer Ausgabegeräte (Bildschirmbrille) und sensorischer Eingabemedien einsteigen und diese manipulieren können. Mit ihrer Hilfe können komplexe Systeme visuell und ansatzweise auch taktil und akustisch erschlossen werden. Neben den Hoffnungen ihrer Proponenten in den Freizeitwert dieser Technik, die bis zur Realisierung einer „visionären" Droge reichen, sind durchaus ernsthafte Anwendungen, zum Beispiel in der Architektur, Medizin, zur Flugsimulation und im Information Retrieval in der Erprobung.

Eine wesentliche Eigenschaft „virtueller Realität" ist die unmittelbare Rückkopplung des Computers mit den menschlichen Aktoren: Sinnesreize führen zu Reaktionen, die in Echtzeit in das System umgesetzt werden, das wiederum in Echtzeit reagiert. In umfassender Weise wird hiermit eine neue Dimension des bekannten „Biofeedback" eröffnet. Durch die unmittelbare Wechselwirkung wird das Beobachtete durch den – aktiven und im System befindlichen – Beobachter verändert, aber auch der Beobachter durch das sich reaktiv verhaltende Beobachtete. Hierbei übernimmt der Computer die Rolle eines Generators von Wirklichkeit, Realität wird quasi „erzeugt aus reiner Vernunft" (Vilem Flusser), ohne materielle Gestalt anzunehmen. Durch die mit dem technischen Fortschritt absehbare immer höher werdende optische und taktile Auflösung wird der wahrnehmbare Abstand zwischen den künstlichen Bildern und dem Bild, das wir uns von unserer Lebenswelt machen, zusehends geringer. Dies stellt Anforderungen an unser Unterscheidungs- und Urteilsvermögen, denen wir vielleicht eines Tages nicht mehr gewachsen sein werden.

Hinter der Entwicklung von Systemen der „virtuellen Realität" steht die Absicht, eine neue qualitative Ebene des Erlebens zu erreichen. Im Unterschied zu herkömmlichen interaktiven Softwaresystemen wie etwa Informations- und CAD-Systemen zielen sie darauf ab, ihre Benutzer in die Lage zu versetzen, „eine Verbindung mit (ihrem) eigenen Intellekt herzustellen oder mit einer virtuellen Gemeinschaft anderer Intellekte" (Leary 1991, S. 280). Damit wird aber auch eine neue Problemqualität erzeugt; es bedarf keiner großen Phantasie, um sich etwa die Gefahr neuer Formen des Autismus (Walkman-Effekt) auszumalen. Das soziale Gefahrenpotential dieser neuen Technik ist bei weitem noch nicht intel-

lektuell aufgearbeitet. *Droht hiermit ein Realitätsverlust durch „Scheinwelten"? Und wer kontrolliert das Vordringen solcher KI-Technologie in den Freizeit- und Bildungsbereich?*

Das Problem des Realitätsverlusts hat jedoch noch eine weitere Facette: Neben einer künstlichen Welt, die im wesentlichen Aspekte unserer Lebenswelt nachahmt und dieser also in hohem Maße ähnlich ist, sind auch künstliche Welten vorstellbar, die sich von ihr radikal unter-scheiden. Welche Auswirkungen solche künstlichen Welten, in denen andere Gesetze gelten und unsere Erwartungen und Erfahrungen, z.B. bezüglich Stabilität und Regelmäßigkeiten der Umwelt, ständig verletzt werden, auf unseren kognitiven Apparat haben, ist unbekannt; es wird notwendig sein, hier ein breit angelegtes Forschungsprogramm in die Wege zu leiten. Die Entwicklung solcher technischen Mittel muß unter das Primat einer Technik nach menschlichem Maß gestellt werden, und das heißt zuerst, daß sie so zu gestalten ist, daß ihre Anwender zu jedem Zeitpunkt über vollständige und bewußte Kontrolle verfügen. Zugleich sind gesellschaftliche Kontrollmechanismen für ihren Entwurf und Einsatz zu entwerfen, so daß „virtuelle Realität" zu einer Technik *für* die Gesellschaft wird und die Destruktion von Individualität und sozialen Werten ausgeschlossen ist.

Wird sich die Disproportionierung der Gesellschaft in „computergläubige Analphabeten" und Menschen mit höherem Bildungsgrad vergrößern? Die zunehmende Computerisierung aller Lebensbereiche droht Menschen mit geringer Bildung, aber auch Gebildete ohne ein Minimum an Computererfahrung sozial auszugrenzen. Es besteht die Gefahr, daß eine relativ große Minderheit von „Rationalisierungsverlierern" und sozial nicht Privilegierten schlicht den Anschluß an eine künftige Gesellschaft verpaßt. Denkbar ist allerdings auch, daß es sich dabei um eine vorübergehende Entwicklung handelt, die durch die Entwicklung intelligenter Mensch-Maschine-Schnittstellen, etwa auf der Basis sprachverstehender Automaten, korrigiert werden kann.

Eine weitere Gefahr liegt in der unreflektierten „Computergläubigkeit" von Benutzern, in deren Augen Sachverhalte und Ergebnisse besonderes Gewicht erhalten, nur weil sie von einem Computer („der kann ja nicht irren") erstellt

worden sind. Bereits heute neigen viele Menschen dazu, Computern menschliche Eigenschaften zuzuschreiben. Diese Entwicklung könnte durch den Einsatz von KI-Systemen, die sich durch die Verarbeitung von Sprach- und Bildinformationen menschlichen Kommunikationsformen annähern, noch forciert werden und dazu führen, daß menschliches Urteilsvermögen hinter suggestiver Autorität von Maschinen zurücktritt.

Fiktive Szenarien: Die sehr abgehobenen Spekulationen von Hans Moravec beziehen sich auf eine Übertragung des menschlichen Geistes auf Computer und somit die Abkopplung vom Körper. Hiermit werden Zukunftsszenarien entworfen, in denen die Fortexistenz menschlichen Kulturguts über die Fortexistenz des Menschen als dominierende Spezies gestellt und der Mensch seiner Vorrangstellung beraubt wird. Es ist schwer, sich Menschen vorzustellen, die sich einem solchen Zukunftsentwurf anschließen wollen. Sich damit extrapolierend auseinanderzusetzen, fällt ebenfalls schwer, da solche Vorstellungen weit außerhalb vergleichbarer Erfahrungsmöglichkeiten des Menschen liegen. Das geeignete Genre dafür ist gegenwärtig am ehesten die Science Fiction.

Fazit

Seit jeher war technologischer Wandel von Ängsten begleitet. Durch ihre Ansprüche, die bis zur Infragestellung des Homo Sapiens in seiner Einzigkeit reichen, schafft Künstliche Intelligenz in einem beträchtlichen Ausmaß Ungewißheit über die Entwicklung einer zukünftigen Gesellschaft. Die Bandbreite publizierter Visionen reicht von Horrorszenarien außer Kontrolle geratener Computer bis zu einem goldenen Zeitalter, wo sinnenfrohe Privatiers alle Unannehmlichkeiten von klugen Robotern erledigen lassen.

Das marktwirtschaftliche Konkurrenzprinzip erzwingt produktionstechnische Rationalität. Der Rechner kam rechtzeitig, um eine zunehmend komplexe Welt handhaben zu können. Künstliche Intelligenz könnte als nächste Stufe dieser Entwicklung betrachtet werden. Das Ergebnis ist ein starker Zuwachs an gesellschaftlicher Komplexität. Die intraindividuelle Kopplung von Funktion und Sinn

zwingt deshalb zur individuellen Reduktion von Umweltkomplexität (vgl. Luhmann 1987, S. 242f). Für die gesellschaftliche Reproduktionsfähigkeit ist wachsende Eigenkomplexität konstitutiv, für Individuen aber die Reduktion von Umweltkomplexität. Aus einer optimistischen Perspektive ließe sich formulieren, daß KI auf lange Sicht ein Mittel ist, die Selbstkonstitution der Menschheit zu fördern, indem sie niedere Formen von Rationalität an Maschinen delegiert.

Diskutiert man die möglichen Auswirkungen von KI und sucht man dabei nach möglichen Parallelen zur Industriellen Revolution, gewinnt die Entkopplung von System und Lebenswelt an Bedeutung. Kennzeichen der Industriellen Revolution war das Delegieren physischer Potenz an Maschinen. Die weitere technische Entwicklung ermöglichte das Delegieren von Koordinations- und Verknüpfungsproblemen auf Maschinen und forcierte die Entkopplung von System und Lebenswelt. Eine wichtige Frage ist, ob künftige Stadien einer entwickelten KI diesen Prozeß des Delegierens menschlicher Fähigkeiten nur graduell fortsetzen, oder ob eine neue Qualität in den Relationen sozialer Systeme entsteht.

Aus einer rein ökonomischen Perspektive sind in diesem Falle die möglichen KI-Folgen mit denen der Industriellen Revolution vergleichbar. Viele Anzeichen sprechen für einen neuen Rationalisierungsschub mit der Folge erhöhter Arbeitsproduktivität und den weiter oben diskutierten Sekundäreffekten. Sofern es dazu kommt, daß „entwickelte KI" in großem Umfang in den Produktionsablauf integriert wird, müßte dies als ein echter Qualitätssprung gewertet werden.

Hiermit verbunden wäre allerdings die Möglichkeit des Wandels des Bildes vom Menschen, insbesondere der Ethik. Ethik erwächst aus der unmittelbaren und individuellen Wahrnehmung konkreten Weltgeschehens; sie verliert sich mit zunehmender Distanz und der Instrumentalisierung der Weltbezüge des Einzelnen. Der wachsende Einsatz von KI wirkt in doppelter Hinsicht auf diese Bezüge ein:

1. Verantwortung gründet sich wesentlich in dem Bewußtsein, Teil einer Gesamtheit zu sein. Bis vor wenigen hundert Jahren bestand der Beitrag des Einzelnen für das Überleben der Gattung hauptsächlich in unmittelbar produktiver Arbeit.

Die Folgen eines individuellen Verantwortungsverlustes bestanden in harten, unmittelbar fühlbaren Sanktionen. Die enge Kopplung von Verantwortung, Arbeit und Sinn ging mit Produktionsweisen einher, die sich erst mit der Industriellen Revolution wandelten. Denkbar ist, daß Verantwortung des einzelnen für eine Sozialgemeinschaft in dem Grade abnimmt, in dem produktive Arbeit und Gattungs-reproduktion entkoppelt werden. Der Kontextverlust des Individuums wird ohne Frage durch zunehmenden Einsatz von KI befördert und begünstigt.

2. Ethik impliziert (noch) eine Sonderstellung des Menschen. Als Subjekte sind ausschließlich Menschen von ethischen Kategorien betroffen, als Objekte genießen sie eine Sonderstellung. Es ist vorstellbar, daß ethische Dimensionen ihren ausschließlichen Bezug auf menschliches Verhalten in dem Maße verlieren, wie Maschinen humane Fähigkeiten simulieren können.

Die ethischen Konsequenzen des KI-Einsatzes sind ambivalent: Künstliche Intelligenz kann Verantwortung zu Fragen der Haftung reduzieren, gleichzeitig aber auch Verantwortung über ein bloßes Arbeitsethos hinausheben. Daß eine menschliche Sonderstellung angefochten wird, ist aber kein einmaliger Vorgang. Im günstigen Fall könnte der KI-Einsatz im Arbeitsbereich jedoch zu einer neuen Selbstkonstitution der Menschheit als Spezies führen, die ihre Zeit nicht hauptsächlich auf Arterhaltung durch Arbeit verwenden muß.

Literatur

Baudrillard, J. (1989). Videowelt und fraktales Subjekt. In Ars Electronica (Hg.): Philosophien der neuen Technologie. Berlin: Merve.

Flusser, V. (1987). Vampyroteutis infernalis. Göttingen: Immatrix Publications.

Leary, T. (1991). Das interpersonale, interaktive, interdimensionale Interface. In M. Waffeneder (Hg.): Cyberspace – Ausflüge in virtuelle Wirklichkeiten (S. 275-281). Reinbek bei Hamburg: rororo Computer (Nr. 8185).

Luhmann, N. (1987). Soziale Systeme. Frankfurt: Suhrkamp.

Moravec, H. (1988) Mind Children. Cambridge (MA): Harvard University Press.

Siekmann, J. (1989). Künstliche Intelligenz: Von den Anfängen in die Zukunft. HMD 150/1989.

Waffeneder, M. (Hg.) (1991). Cyberspace – Ausflüge in virtuelle Wirklichkeiten. Reinbek bei Hamburg: rororo Computer (Nr. 8185).

KI-Produkte und Verantwortung

P. Schreiber

Wer ist verantwortlich für die Ergebnisse, die von KI-Maschinen „erarbeitet" werden? Wie sieht die Interaktions- und Verantwortungsstruktur aus bei einer neuen Rollenverteilung zwischen Mensch und Maschine? Der Aussage „Maschinen handeln nicht", steht der Satz gegenüber: „Das ist eine neue Art von Maschinen, die uns Bedingungen stellen". Dazu ein Beispiel aus der Fertigungstechnik: Eine *Maschine* mit *KI-Eigenschaften* könnte auch ein Netzwerk aus eigenständigen Rechnern sein. Zum Beispiel ein System mit CAD-Büro, Arbeitsvorbereitung und Fertigungshalle. KI hat nur Sinn, wenn die Rechner Routineaufgaben und „etwas" darüber hinaus autonom erledigen. Das heißt, daß möglicherweise 60 % der Rechnertätigkeit bzw. -entscheidungen ohne menschliches Einwirken ablaufen werden. Daraus erwächst die Aufgabe, *Überwachungsalgorithmen* zu entwickeln, die einmal die Qualität der Arbeiten sicherstellen, zum anderen bei Gefahren beinhaltenden Tätigkeiten weit verantwortungsvollere Aufgaben zu erledigen haben.

Verantwortung ist ein Begriff aus dem philosophischen Sprachraum, der Ethik. Wir unterscheiden individuelle und kollektive Verantwortung. Nach Arthur Schopenhauer bedeutet Verantwortlichkeit *die Möglichkeit, anders gehandelt zu haben*. Im Gegensatz zu der bis heute bekannten, voll determinierten Maschine hat aber ein KI-Produkt – wenn es diesen Namen verdient – gerade diese „Möglichkeit, anders gehandelt zu haben". Daraus ist ersichtlich, daß hier einiges unscharf werden wird. Es ist eine weitere (geistige) Revolution – oder eine weitere kopernikanische Wende. Ein KI-Produkt, das den Turingtest besteht, ist als solches bei einem „verdeckten" Dialog – z. B. über ein Terminal – nicht zu erkennen.

Deswegen werden solche Systeme nicht zu *Personen*, aber es sind *Akteure* für eine komplexe Aufgabe; ausgestattet mit den dazu erforderlichen „Freiheiten".

Es ist schon ein Einbruch in die bisher Personen vorbehaltene Domäne „Bewerten, Entscheiden, Verantworten" bei eingeschränkten rationalen Sachverhalten, und es stellen sich z. B. die folgenden Fragen (die hier nicht beantwortet werden können):

- Ist mit der Kompetenzzuweisung auf Maschinen eine Reduktion ethischen Verhaltens verbunden?
- Geht damit eine Aushöhlung von individueller Verantwortung einher?

Im täglichen Leben tritt an Stelle der Verantwortung die *Haftung*. Aber die Begriffe sind nicht austauschbar. Verantwortung beinhaltet z. B. auch die gesamten prophylaktischen Maßnahmen (verantwortungsvolles Handeln). Haftung wirkt überwiegend erst, nachdem das eigentlich zu verhindernde Ereignis eingetreten ist, und verarmt letztendlich fast immer zu einer rein pekuniären Angelegenheit. Deshalb wird der Begriff „Verantwortung" in einem mehr praktischen Sinne hier weiter verwendet.

Dies alles bedeutet für die Jurisprudenz bez. der *Verantwortlichkeit von Maschinen* einige schwer zu lösenden Fragen. Die bisher einfache Unterscheidung in Menschen und Sachen könnte schwierig werden und muß differenzierteren Betrachtungen weichen.

Neue Aufgaben wird es auch für die Polizei geben. Wer hätte sich zur Zeit, als die ersten Benzkutschen über Pflasterstraßen rollten, vorstellen können, daß es einer eigenen starken Verkehrspolizei bedarf. Einen ähnlichen, aber viel schnelleren Aufschwung wird die *Datenpolizei* nehmen. Es wird KI-Automaten geben, betreiberfern in Netzen eingebunden, geübt in der Manipulation von Bankkonten und in raffiniertesten Verschleierungstaktiken (Datenvernichtung, Fehlfährten), die nur die Aufgabe haben, das Wohl eines einzelnen oder einer Clique zu mehren und die dabei ihr ganzes nichtdeterministisches Potential einsetzen, damit der Betreiber nicht gefunden wird. Diese Mißbräuche gilt es zu unterbinden. Das eigentliche „idealistische" Ziel der Arbeiten zur Entwicklung der KI – wie auch zu anderen Zukunftstechnologien – sollte eine verbesserte Lebensqualität für die Menschen und die Zukunftssicherung der Erde sein . Das heißt, die KI muß

bemüht sein, Systeme bereitzustellen, mit denen solche Aufgaben unterstützt werden.

Im einzelnen einige Beispiele:

- *Leistungsfähigkeit.* Dieser Aspekt wird von niemandem ignoriert. Besonders aufmerksam gemacht werden muß aber auf andere Aspekte:
- *Transparenz und Partizipation.* Die Entwurfsziele von KI-Systemen sollten für die vom Einsatz solcher Systeme Betroffenen in verständlicher Weise offengelegt und die Systeme unter Beteiligung entwickelt werden.
- *Verläßlichkeit und Vertraulichkeit.* Wie alle Technik ist auch KI nicht fehlerfrei; aus diesen und anderen später dargelegten Gründen gilt es, für die Überwachung solcher Systeme Sorge zu tragen (Plausibilitätskontrollen, Erklärungskomponenten. Weiterhin ist sicherzustellen, daß der Einsatz von KI-Systemen nicht das Grundrecht auf „informationelle Selbstbestimmung" verletzt.

Dies sind ethische Postulate, die eine entsprechende Technikbewertung erfordern, inwieweit sie zu erreichen sind und an welcher Stelle die Gefahren lauern. Rapp (1989) beschreibt die Schwierigkeiten der Technikbewertung und der Umsetzung der Ergebnisse derselben in die Praxis. Zwei seien hier genannt:

- Bedürfnisexplosion – erreichte Ziele wecken neue Erwartungen – und
- begrenzte Fähigkeiten zur theoretischen Erfassung des wissenschaftlich-technischen Wandels und zur Prognose.

Die Frage der *ethischen* und *gesellschaftlichen Verantwortbarkeit* der KI stellt sich also nicht nur bei „kriminellen Akteuren", sondern auch z. B. bei Expertensystemen, die mit ihren Ergebnissen eine Gefahr für einzelne (z.B. beim Einsatz im medizinischen Dienst) oder für viele (bei Arbeits- und Umweltschutzrelevanz) Menschen heraufbeschwören können. Dies zeigt, daß ein „selbständig schließendes" KI-Programm anders zu bewerten ist als ein Fachbuch, in dem sich möglicherweise ein Fehler befindet. Letztendlich werden dies immer im einzelnen zu behandelnde Fragestellungen sein, weil KI-Maschinen viele Facetten neben dem Rationalisierungspotential und möglichen negativen Aspekten bein-

halten können, z.B. Befreiung von gesundheitsgefährdenden Tätigkeiten oder Schaffung einer größeren Flexibilität der Arbeit (z.B. Tele-Heimberufe), und nicht zuletzt die leichtere Überwindung von Behinderungen, insbesondere wenn diese die Sinnesorgane betreffen.

Ein weiterer Problemkreis ergibt sich aus den möglichen *Anteilen für eine Haftung*, bezogen auf den Hersteller des KI-Produktes, auf den Hersteller z.B. einer fertigungstechnischen Anlage und auf den Betreiber des Gesamtsystems. Die klassische *Produkthaftung* des Herstellers setzt im Prinzip die Übergabe einer voll determinierten Anlage oder Maschine voraus, in der jeder (logische) Fehler – soweit es sich nicht um *Zufallsfehler* handelt (z.B. Ausfall eines Chips) – von Anfang an vorhanden ist *(systematischer Fehler)*. Bei einem KI-Produkt ist dies per definitionem unmöglich. Es lernt (erweitert die Wissensbasis) und schließt autonom. Das heißt, es unterliegt beim Betreiber Veränderungen, auf die der Hersteller keinen Einfluß hat. Andererseits sind diese Veränderungen in gewisser Weise doch vom Hersteller durch die Art beeinflußt, wie er das Programm konzipiert hat. All dies gilt auch für abgrenzbare Problemstellungen. Klassische Lösungen, wie die heutigen Analysen oder TÜV-Abnahmen, sind nur beschränkt anwendbar, da die (selbständige) Weiterentwicklung beim Betreiber nicht vorausgesehen werden kann. KI-Produkte sind keine abgeschlossenen Systeme, die zumindest grundsätzlich mit allen Implikationen vollständig testbar wären; abgesehen davon, daß diese KI-Programme von Hause aus umfänglich und komplex sein werden. Fragen der Anlagensicherheit (im Sinne der Integrität der Anlage), des Arbeits- und des Umweltschutzes bedürfen demnach *besonderer Analysen* und *zusätzlicher Absicherungen*, wenn der „Freiheitsgrad" des KI-Produktes diese mit Vorrang sicherzustellenden Schutzanforderungen tangiert.

Die Überlegungen zeigen, daß hier Elemente vorliegen, die in Richtung auf eine *kollektive* Verantwortung (Haftung) weisen, obwohl dies in der Praxis aus Gründen der Rechtssicherheit möglichst vermieden wird. Es ist denkbar, daß Unsicherheiten in diesen Fragen sich als ein wesentlicher Hemmschuh bei der Diffusion von KI-Produkten entsprechenden Aufgabenzuschnitts auswirken wird.

Angenommen, die Jurisprudenz weist dem Betreiber des Gesamtsystems die Verantwortung zu, dann kann er diese nur übernehmen, wenn er vom Hersteller des KI-Produktes (Systems) eine entsprechende Unterstützung erfährt. In der Praxis heißt dies, daß entsprechend den Ergebnissen der Analysen möglicher Gefahren vom Hersteller bei solchen KI-Produkten *prophylaktisch wirkende Verantwortungsfunktionen* – wie sie nachstehend an zwei Beispielen näher beschrieben werden – vollständig angelegt sein müssen (Verantwortung im praktischen Sinne; Haftung assoziiert zu sehr mit nachträglicher kurativer oder geldlicher Entschädigung).

Diese „Verantwortungsfunktionen" beziehen sich nur auf die Aufgaben, die ein solches KI-Produkt hat (z. B. eine Motorendiagnose), und setzen (weitgehend) *Fehlerfreiheit in Hard- und Software* voraus. Das heißt, daß alles, was zu diesem Zweck an Maßnahmen durchgeführt wird (z. B. Fehlertoleranz, Tests, diversitäre Softwarealgorithmen etc.), unterhalb der Ebene der so definierten „Verantwortungsfunktionen" angesiedelt ist. Dies darf nicht miteinander vermischt bzw. verwechselt werden. Auch würden sich solche „Verantwortungsfunktionen" nicht auf KI-Produkte mit sicherheitsrelevanten Anwendungen beschränken. Ein Spekulationsprogramm für die Börse muß auch als KI-Produkt „geordnet" funktionieren, sonst gibt es Ärger für den Betreiber und den Hersteller. Hier stehen allerdings überwiegend Aufgaben zur Diskussion, die eine entsprechende Analyse von Fragen der Verantwortlichkeit erfordern.

Allgemeiner faßt Gatzemeier (1990) die Bedeutung von „Verantwortung" für ein Industrieunternehmen und seinem Personal in sieben Thesen zusammen, zum Beispiel:

- Angemessene Berücksichtigung der Folgen des Handelns für alle potentiell Betroffenen,
- eine steigende Anzahl der Betroffenen und Qualität der Folgen verpflichtet alle zu erhöhter Sensibilität gegenüber möglichen Entscheidungsfolgen und
- die Verantwortung selbst ist unteilbar. Die Wahrnehmung der Verantwortung ist nach den Einflußmöglichkeiten zu differenzieren.

Diese in den Betrieb hineingerichteten Überlegungen müssen vielleicht nur wenig geändert werden, wenn eine „kollektive" Verantwortung für Hersteller und Betreiber bei KI-Produkten postuliert wird, aber zu den Beeinflussungsgraden und -möglichkeiten in bezug auf „das Handeln dieser neuartigen Akteure" sind noch viele Untersuchungen erforderlich.

Zurück wieder zu den KI-Systemen selbst, wäre es verfrüht, jetzt schon solche *Verantwortungsfunktionen* näher charakterisieren zu wollen. Dies muß eine spezielle Forschung erbringen, die von der Materie her teilweise auch außerhalb des engeren KI-Feldes anzusiedeln ist. Die Diskussionen und ersten Erfahrungen mit (noch sehr unausgereiften) KI-Produkten zeigen, welche Funktionen – möglichst in Form von (fast) unabhängigen Softwaremoduln – in enger Beziehung zum Problemkreis „Verantwortlichkeit" von KI-Produkten stehen. Zwei seien als Beispiele angeführt:

Eine dieser notwendigen Funktionen ist die separate, zusätzliche Überwachung, ob ein KI-Produkt sich noch im Rahmen seiner Kompetenz bewegt. Diese *Kompetenzüberwacher* sollen eine Art „Gewissen" sein, das sich meldet, wenn Schlußweisen oder die Schlüsse selbst (bei Expertensystemen oder Bildverarbeitung) oder Handlungen (bei Robotern oder Anlagen jeder Art) nicht vom fachlichen Kompetenzbereich des KI-Systems her abgedeckt sind. Diese Überwachungsfunktionen sollen die KI-Produkten nachgesagte Blindheit für die Grenzen ihres Kompetenzbereiches (Stammtischverhalten) überwinden helfen (Moldaschl 1990).

Eine weitere „Verantwortungsfunktion" ist notwendig für die Unterstützung des Menschen bei der *Überwachung der Ergebnisse* eines KI-Systems. Am Anfang dieses Abschnitts wurde ein (zukünftiges) vernetztes KI-System, bestehend aus CAD-Büro, Arbeitsvorbereitung und Fertigungshalle, angesprochen. Ein Großteil des Informationsaustauschs und der Entscheidungsfindung läuft ohne Eingriff durch eine Person ab (alle Routineaufgaben und einiges darüber hinaus). Darunter werden sich auch solche Kalkulationen befinden, die in Handlungen übertragen werden, die für die Mitarbeiter arbeitsschutz- bzw. gesundheitsschutzrelevant sind bzw. werden können. Diese Überlegungen lassen sich auf

Produktqualität, Anlagensicherheit und Umweltschutz gleichermaßen übertragen. Je nach dem zu erwarteten Risiko muß hier eine abgestufte Überwachung erfolgen.

Die entsprechenden „Verantwortungsfunktionen" seien hier *Automatenüberwacher* genannt. Dies können keine individuellen Realisierungsversuche sein. Dazu wird eine allgemeine Theorie zur „Überwachung von selbständig schließenden Automaten" benötigt, von der *Leitfäden* für die Vorgehensweise im Einzelfall abgeleitet werden können. Anhand von Kriterien aus der Aufgabe, der Anlage und des KI-Systems und den damit verbundenen Risiken müßte beispielsweise das „was" (an Information) und das „wie oft" unter Einbeziehung praktischer Erfahrungen abgeleitet werden können, für die Einbindung des überwachenden Menschen in diesen Prozeß.

Eine davon zu trennende Aufgabe der *Automatenüberwacher* ist die „Form", in der die zur Kontrolle notwendigen Daten angeboten werden, das „wie". Erstens ist eine Aufbereitung in Form von „Erklärungsfunktionen" notwendig, die Sachverhalte und Hintergründe für die Schlüsse des KI-Systems aufzeigen, statt dem Menschen den Nachvollzug von Speichermanipulationen zuzumuten. Zweitens sollten sequentiell ablaufende Vorgänge bei den Rechnern und im Arbeitsablauf in einer Form umgesetzt werden, die ein gleichzeitiges Erfassen des Wesentlichen erlauben. In einem Bild, mit vergleichenden Bildern (Soll/Ist) oder einer Bilderfolge, können z. B. Unstimmigkeiten mit relativ geringem Aufwand von Menschen schnell erfaßt werden. Die Erfolge der graphischen Benutzeroberflächen sprechen für sich.

Benötigt eine solche „Verantwortungsfunktion" aus Sicherheitsgründen zusätzlich die Überwachung durch eine oder mehrere Personen, dann muß die dazu erforderliche *Mensch-Maschine-(MM-)Schnittstelle* ihre Mittlerfunktionen unter allen Umständen wahrnehmen können. Das heißt, daß alle Aspekte, die mit der notwendigen Übernahme der „Verantwortung" durch einen Menschen zusammenhängen, konstruktiv berücksichtigt sind und diese Vorkehrungen ein entsprechendes Zuverlässigkeitsniveau haben. Für einen größeren autonomen Mobilroboter bedeutet dies – um nur ein sehr einfaches Beispiel zu nennen –,

daß er bei einer falschen Entscheidung und insbesondere im gestörten Zustand sicher abgeschaltet werden kann; z. B. mittels Not-Aus über Funkbefehl.

Inwieweit hier eine unter allen Umständen sichere, eine fehlertolerante oder nur eine normal zuverlässige MM-Schnittstelle eingesetzt werden muß, hängt vom Risiko ab, das bei Fehlern (falsches Ergebnis) oder Fehlfunktionen des KI-Systems zu besorgen ist. Eine Folge könnte sein, daß sich die Notwendigkeit einer noch schärferen Trennung der Programmteile für die „Benutzeroberfläche" (MM-Schnittstelle) und für die eigentliche „Fachaufgabe" ergibt, als sie heute üblich ist. Beide können prinzipiell KI-Produkte sein.

Dies sind (nur) zwei Beispiele (Kompetenz- und Automatenüberwacher) von Aufgaben, die von *Verantwortungsfunktionen* wahrzunehmen sind.

Man kann sich vorstellen, daß die Anzahl unterschiedlicher „Verantwortungsfunktionen" auf eine bestimmte Zahl von *Grundtypen* zu begrenzen ist. Wenn dem so ist, dann sollten dazu *Standardlösungen* (Hard- und Software) entwickelt werden können, die leichter geprüft (und später normiert; gerichtliche Nachprüfung) werden können, als wenn jeweils zu jedem sicherheitsrelevanten KI-Produkt individuell solche Funktionen erarbeitet werden und dann jeweils auch so geprüft werden müssen. KI-Systeme sind dann vielleicht „normale" Produkte geworden und aus dem Schutzwall der Forschung entlassen.

Dies alles ist leicht hingesagt. Wichtig für die Zukunft ist folgendes: Eine spezielle Forschung zu diesen *Verantwortungsfunktionen* in KI-Produkten, die fast immer benötigt werden (Produkthaftung) und bei höheren Sicherheitsanforderungen nur entsprechend effektiver sein müssen, sollte aufzeigen, was an diesen Funktionen verallgemeinerbar und in Regeln, in Abhängigkeit von Parametern, formuliert werden kann und was grundsätzlich für jedes KI-Produkt mit entsprechenden Anforderungen jeweils neu erarbeitet werden muß. Die Arbeiten zu dieser Art von Sicherheitsforschung, die zum Teil wenig gemein hat mit der Facharbeit von Informatikern, sollte parallel zur KI-Weiterentwicklung durchgeführt werden und einen wichtigen Beitrag zur Einführung von KI-Produkten leisten, damit Menschen dabei durch eine falsche Einschätzung „vermeintlicher

Sicherheiten" bei der Lösung der jeweiligen Aufgabe durch Produkte einer fortschrittlichen Informationstechnologie – persönlich und materiell – nicht zu Schaden kommen.

Literatur

Gatzemeier, M. (1990). Was bedeutet Verantwortung für ein Industrieunternehmen und seine Mitarbeiter/innen. In: K. Henning, A. Bitzer (Hrsg.), Ethische Aspekte von Wirtschaft und Arbeit. Mannheim: BI-Wissenschaftsverlag.

Moldaschl, M. (1990). Das Modell ist gut, nur die Realität ist schlecht. Technische Rundschau 49/90, S. 104.

Rapp, F. (1989). Technischer Wandel und ethische Postulate. In: M. Gatzemeier (Hrsg.), Verantwortung in Wissenschaft und Technik. Mannheim: BI-Wissenschaftsverlag.

Potentielle Gefahren von KI-Systemen

H. Röpke

Bekanntlich ist die Verwendung von Werkzeugen – wie auch der Einsatz von KI-Systemen – ambivalent, d.h. sie können sowohl Vorteile als auch Nachteile bewirken. Das gilt für sachliche Auswirkungen ebenso wie für die im sozialen Bereich. Sieht man von einem zielgerichteten negativen Einsatz der KI ab, so gibt es dennoch unbeabsichtigte negative Nebenwirkungen, die selbst bei vernünftigen und sachgerechten Anwendungen eintreten können. Deshalb ist es wichtig, die potentiellen Gefahren zu erkennen und zu versuchen, sie durch vorbeugende Maßnahmen aufzufangen.

Potentielle Gefahren sind relativ leicht zu nennen, welche davon aber wirklich eintreten werden, ist natürlich ungewiß. Geht man jedoch davon aus, daß KI-Systeme aus der Sicht der Datenverarbeitung die gravierendsten Umwälzungen im geistigen Bereich bewirken können, so dürften sich vor allem die gesellschaftlichen Entwicklungen verstärken, die man schon seit Beginn der maschinellen Datenverarbeitung tendenziell beobachten kann, z.B. die stärkere Schematisierung beruflicher Abläufe.

Selbstverständlich hängen gesellschaftliche Veränderungen von der Grundeinstellung der Menschen ab, die ihrerseits wieder von der politischen Zielsetzung des Staates beeinflußt werden kann. Da moderne Industriegesellschaften – wie die der Bundesrepublik Deutschland – die KI als förderungswürdige Schlüsseltechnologie betrachten, wird sie entsprechend subventioniert. Dadurch ist die KI automatisch in den strukturpolitischen internationalen Wettlauf eingebunden, was ihrer selbstkritischen Betrachtung nicht förderlich ist.

Durch den hohen Anspruch der KI-Systeme und des werbewirksamen Auftretens ihrer Fürsprecher sind extrem große Erwartungen geweckt worden. Beim Eindringen der KI in die Bereiche der Arbeitswelt werden aber negative Auswirkun-

gen nicht zu vermeiden sein. Einige potentielle Gefahren sollen im folgenden kurz umrissen werden.

Computergläubigkeit

Durch den verstärkten KI-Einsatz wird vermutlich – mindestens bei einem Teil der Gesellschaft – die „Computergläubigkeit" zunehmen. Es ist nicht nur verlokkend, dem Computerausdruck blind zu vertrauen, sondern darüber hinaus auch unmöglich, die Ergebnisse von komplexen Problemlösungsprozessen vollständig und zuverlässig zu überprüfen. Auf lange Sicht kann das bei den Nutzern solcher von vielen Meinungsbildnern gepriesenen KI-Systeme zur Kritiklosigkeit führen. So könnten z.B. Ergebnisse prinzipiell akzeptiert und Alternativen weder gesucht noch berücksichtigt werden, so daß die Verantwortung für die Ergebnisse nach dem Selbstverständnis dieser unkritischen KI-Systembenutzer auf die Technik (oder den Computer) übertragen wird.

Wegen der großen Verarbeitungsgeschwindigkeit der Digitalrechner und ihrer technisch perfekt erscheinenden Arbeitsweise kann für die Nutzer der KI-Systeme durchaus der Eindruck entstehen, daß die „künstliche" Intelligenz etwas Eigenständiges, eine vom Menschen losgelöste „höhere" Form von Intelligenz ist, an die man unverrückbar glauben darf oder muß! Dieser Eindruck wäre verhängnisvoll, und es ist daher notwendig, immer wieder darauf hinzuweisen, daß man mit „Künstlicher Intelligenz" wissensbasierte Programmsysteme bezeichnet, die mit menschlicher (d.h."natürlicher") Intelligenz entwickelt und auf Maschinen gespeichert worden sind, so daß sie aufgrund ihrer Vorherbestimmtheit wohldefinierte Probleme lösen können.

Computer-Dialog

Schon jetzt gibt es viele Computer-Nutzer, die von einem „Dialog" zwischen Mensch und Maschine sprechen, obwohl auch eine noch so „intelligent" programmierte Maschine keinen Dialog führen kann. Diese ebenso saloppe, wie irreführende Ausdrucksweise wird wahrscheinlich durch den verstärkten Einsatz

von KI-Systemen häufiger benutzt und sich vielleicht sogar etablieren. Man wird noch öfter als bisher von einem „Partner", also von einem menschenähnlichen Wesen sprechen. Diese Wortwahl „Mensch/Maschine-Dialog" ist aber nicht nur sprachprägend, sondern auf die Dauer auch meinungsbildend! In den Augen vieler Menschen ist der Computer ja heute schon wenn nicht ein „richtiger" Mensch, so doch eben ein „selbständiger Akteur". Er ist schließlich schon lange ein ausgezeichneter „Schachspieler", den nur wenige noch zu besiegen imstande sind! Trotzdem oder gerade deshalb muß man immer wieder vor dieser leichtfertigen „Vermenschlichung" technischer Systeme warnen. Zur Dialogfähigkeit gehört die Reflektionsfähigkeit, die auch intelligent programmierte Maschinen und KI-Systeme nicht besitzen. Es fehlt ihnen an Alltagswissen, Bewußtsein, Willenskraft, Initiative, Kreativität, Eigeninteresse, Intuition usw. Man sollte sich deshalb immer vergegenwärtigen, daß auch ein hochkomplexes KI-System hierbei keine Ausnahmestellung einnimmt. Und deshalb kann ihm auch keine dem Menschen vergleichbare Verantwortung zugesprochen werden, wie es mancherorts schon geschehen ist.

Sprachverarmung

Zu befürchten ist auch eine negative Veränderung des Sprachgebrauchs. Die schon kritisierte, saloppe, vielfach falsche und mitunter auch sehr dürftige Ausdrucksweise weiter Kreise innerhalb der Informatik könnte sich vor allem in denjenigen Arbeitsbereichen weiter verschlechtern, in denen KI-Systeme Sprachübersetzungen vornehmen. Aus verschiedenen Gründen müssen bei der maschinellen Sprachübersetzung viele Standardformulierungen benutzt werden. Sofern sie sachlich vertretbar und einigermaßen verständlich sind, werden sie wahrscheinlich – nach schneller Gewöhnung – auch den eigenen Umgang mit der deutschen Sprache beeinflussen, so daß ein Verlust an Ausdrucksvielfalt eintreten könnte. Da eine Überarbeitung der Texte von maschinellen Übersetzungen aufwendig und nur durch Überwindung der eigenen Bequemlichkeit möglich ist, wird man in der Regel auf Korrekturen verzichten. Die Folge davon wäre ein Abgleiten in die Monotonie der maschinellen Sprache, also eine beklagenswerte Verarmung der Ausdrucksweise. Ebenso beklagenswert wäre eine weitergehende

Disproportionierung der Gesellschaft in „Computerkundige" und „Computer-Analphabeten", deren gesellschaftliche Auswirkungen nicht absehbar sind. Schließlich ist auch zu befürchten, daß eine Verkümmerung von menschlichen Fähigkeiten eintreten wird, die durch eine allgemeine Verfügbarkeit von KI-Systemen nicht mehr benötigt werden. So kann man schon heute immer häufiger beobachten, daß viele Menschen lieber mehrere Minuten Zahlen in ihren Taschenrechner eintippen, als sie in unvergleichbar kürzerer Zeit durch „Kopfrechnen" zu addieren.

Fehlentwicklungen

Angesichts der großen Anstrengungen der Universitäten im KI-Bereich kann davon ausgegangen werden, daß schon in kurzer Zeit einsatzfähige KI-Systeme auf den Markt gelangen werden. Großangelegte Werbungen von Softwarefirmen werden das übrige tun, so daß es vielleicht bald zum guten Ton gehören wird, sich schon aus Prestigegründen dieser Systeme zu bedienen. In einem solchen Entwicklungsstadium werden erfahrungsgemäß die sonst üblichen Kosten/Nutzen-Überlegungen in den Hintergrund gedrängt. Das wiederum ist dann die große Stunde der Nachahmer und unseriösen Anbieter, die andererseits (wegen der Unmöglichkeit von Validierungen hochkomplexer Systeme) vor einer schnellen Entlarvung sicher sein können.

Außer dieser „Verführung" durch Werbung und bewußte Täuschung gibt es noch den Mißbrauch der epistemischen Autorität, der beim Einsatz wissensbasierter Systeme häufiger als anderswo auftreten dürfte. (So z.B. durch den Verweis auf literaturbekannte Namen von Wissenschaftlern, deren Wissen in dem betreffenden KI-System eingeflossen ist.) Problematisch bleibt auch, wie die Wissensbasis auf dem jeweiligen aktuellen Stand des Wissens gehalten werden kann. Sollten derartige Fehlentwicklungen oder Probleme in verstärktem Umfang eintreten, so ist der Gesetzgeber aufgerufen, ihnen durch Präzisierung, Modifizierung oder Ergänzung der betreffenden Gesetze Einhalt zu gebieten.

Ängste und Risiken

Neuerungen erwecken oft Ängste bei den Menschen, deren Arbeit davon stark betroffen ist, und zwar selbst dann, wenn ihre Arbeit dadurch erleichtert oder verbessert werden soll. Das dürfte im Prinzip auch für die KI-Systeme gelten, denn einige der davon Betroffenen werden durch sie ihren Arbeitsplatz gefährdet sehen („Computer als Jobkiller"), und andere wiederum befürchten, den veränderten Bedingungen nicht gewachsen zu sein. Hier besteht aber durchaus die Chance, durch rechtzeitige Ausbildung und Vorbereitung auf die neue Situation oder auch durch Schaffung neuer Beschäftigungsfelder die kommende Problematik zu entschärfen. Bedauerlich wäre es dagegen, wenn aufgrund derartiger Versäumnisse die KI als Vehikel für den arbeitspolitischen Kampf zwischen Gewerkschaften und Arbeitgeberverbänden benutzt würde.

Bei technischen Neuerungen wird des öfteren davon gesprochen, daß der „Störfaktor Mensch" ausgeschaltet werden müsse, da menschliches Versagen oftmals eine Hauptursache von Katastrophen sei. Es kann zwar keinen Zweifel daran geben, daß genügend ausgereifte Systeme in vielen Fällen die Sicherheit von technischen Geräten und Großanlagen beträchtlich erhöhen. Dennoch darf diese Tatsache aber nicht darüber hinwegtäuschen, daß auch bei der Entwicklung von KI-Systemen Fehler möglich sind, die verheerende Auswirkungen haben können. Eine vorschnelle Akzeptanz von Neuerungen führt oft zu größeren Risiken, wenn nicht zuvor gründlich über die Folgen nachgedacht wurde. So trivial die Aussage auch sein mag: Auch beim Einsatz noch so perfekt erscheinender KI-Systeme gibt es kein Null-Risiko! Deshalb kann man nur hoffen, daß bei großflächiger Einführung der KI-Systeme verantwortungsvoll und besonnen vorgegangen wird.

Mit dem Vordringen neuer Technologien sind komplexe gesellschaftliche Wechselwirkungen verbunden. Viele Ereignisse beeinflussen sich gegenseitig, und es ist deshalb schwierig, zuverlässige Vorhersagen über die tatsächlichen Folgen einer einzigen Komponente – wie z.B. die Einführung von KI-Systemen – zu machen. Jede Abschätzung muß also mit Vorbehalten durchgeführt werden, sofern sie nicht nur reine Spekulation bleiben soll. Sehr wesentlich ist natürlich

auch die Geschwindigkeit, mit der arbeitstechnische Veränderungen vorgenommen werden. Geht man in angemessenem Tempo voran, so können unerwünschte Nebeneffekte sicherlich durch flankierende Maßnahmen gemäßigt werden, während unbekümmertes schnelles „Vorwärts-Streben" schlimme Konsequenzen zur Folge haben kann.

Wenn zuvor Sinnfragen gestellt und erörtert werden, die Kritikfähigkeit geschärft und für ausreichende Rückkopplung im sozialen Bereich gesorgt wird, können Ängste und Risiken abgebaut werden, ohne dabei auf die positiven Aspekte technischer Neuerungen verzichten zu müssen.

Auch KI-Systeme sind nur Werkzeuge der Menschen. Sie dürfen nicht mystifiziert, sondern sollten innerhalb verantwortbarer Grenzen eingesetzt werden, um der Gesellschaft als Ganzem zu dienen.

Anhang

Autoren bzw. Mitglieder des VDI-Ausschusses „Künstliche Intelligenz“

Prof. Dr. Armin B. Cremers (Obmann)	Rheinische Friedrich-Wilhelms-Univ. Bonn Institut für Informatik III Römerstr. 164 5300 Bonn 1
Dr. Barbara Becker	GMD Birlinghoven Institut für angewandte Informationstechnik Forschungsgruppr Expertensysteme Postfach 1240 5205 St. Augustin 1
Prof. Dr. Rafael Capurro	Fachhochschule für Druck Feuerbacher Heide 38–42 7000 Stuttgart
Prof. Dr. Rolf Eckmiller	Universität Düsseldorf Institut für Physikalische Biologie Abt. Biokybernetik Universitätsstr. 1 4000 Düsseldorf
Prof. Dr. Günther Görz	Universität Erlangen-Nürnberg IMMD 8 – KI Am Weichselgarten 9 8520 Erlangen
Rolf Haberbeck M.A.	SNI Gustav-Meyer-Allee 1 1000 Berlin 65

Prof. Dr. Andreas Kemmerling	Ludwig-Maximilans-Universität München Institut für Statistik Geschwister-Scholl-Platz 1 8000 München 22
Prof. Dr. Sybille Krämer	Freie Universität Berlin Institut für Philosophie Habelschwerdter Allee 30 1000 Berlin 33
Dr. Anton Kremeier	Staatliches Materialprüfungsamt NRW Marsbruchstr. 186 4600 Dortmund 1
Dr. Rolf A. Müller	Daimler Benz AG Ressort Forschung und Technik Forschungsinstitut Berlin Alt-Moabit 91 b 1000 Berlin 21
Dr. Horst Röpke	Schering AG Postfach 65 03 11 1000 Berlin 65
Dr. Jürgen Seetzen	VDI/VDE-Technologiezentrum für Informationstechnik GmbH Budapester Str. 40 1000 Berlin 30
Prof. Dr. Andreas Schlachetzki	Technische Universität Braunschweig Institut für Halbleitertechnik Postfach 33 29 3300 Braunschweig
Dr. Paul Schreiber	Bundesanstalt für Arbeitsschutz Vogelpothsweg 50–52 4600 Dortmund

Prof. Dr. Gerhard Strube	Albert-Ludwigs-Universität Freiburg Institut für Informatik und Gesellschaft Abteilung Kognitionswissenschaft Friedrichstr. 50 7800 Freiburg i. Br.
Prof. Dr. Ipke Wachsmuth	Universität Bielefeld Technische Fakultät Wissensbasierte Systeme Postfach 86 40 4800 Bielefeld 1
Prof. Dr. Karl F. Wender	Universität Trier Fachbereich 1 Psychologie Postfach 38 25 5500 Trier

Wissenschaftliche Mitarbeiter

Matthias Butt	Bundesallee 90 1000 Berlin 41
Dipl.-Soz. Michael Wilker	Am Osterberg 14 4515 Bad Essen 1
Gernot Grube	Bautzener Str. 17 1000 Berlin 62

Aktuelle Veröffentlichungen aus der VDI-Hauptgruppe

Die folgenden Veröffentlichungen sind über die VDI-Hauptgruppe bzw. über den VDI-Verlag Vertriebsleitung, Postfach 101054, 40001 Düsseldorf zu erhalten.

Kurt A. Detzer: Unsere Verantwortung für eine umweltverträgliche Technikgestaltung. Von abstrakten Leitsätzen zu konkreten Leitbildern. (VDI-Report Nr. 19, 1993) (Hrsg.: VDI-Hauptgruppe)

Kurt A. Detzer: Von den zehn Geboten zu Verhaltenskodizes für Manager und Ingenieure. Was sagen uns ethische Leitsätze, Prinzipien und Normen? (VDI-Report Nr. 11, 3. erg. und überarb. Auflage; 1992 (Hrsg.: VDI-Hauptgruppe)

Empfehlung des VDI zur Integration fachübergreifender Studieninhalte in das Ingenieurstudium, Juli 1990 (Hrsg.: VDI-Hauptgruppe)

Hubert Gräfen (Hrsg.): Die fachübergreifende Qualifikation des Ingenieurs: Anforderungen der Wirtschaft - Angebote der Hochschulen. Düsseldorf (VDI-Verlag) 1990

Handlungsempfehlung: Sozialverträgliche Gestaltung von Automatisierungsvorhaben, Dezember 1989 (Hrsg.: VDI-Hauptgruppe)

Ingenieurverantwortung und Technikethik. Standpunkte - Informationen - Aktivitäten. Broschüre der VDI-Hauptgruppe Düsseldorf 1991

Integrierter Umweltschutz. Ingenieurkonzepte für eine umweltverträgliche Technikgestaltung. VDI-Bericht Nr. 899. Düsseldorf (VDI-Verlag) 1991

Friedrich Rapp; Manfred Mai (Hrsg.): Institutionen der Technikbewertung. Standpunkte aus Wissenschaft, Politik und Wirtschaft. Düsseldorf (VDI-Verlag) 1989

Walther Ch. Zimmerli (Hrsg.): Herausforderung der Gesellschaft durch den technischen Wandel. Informationstechnologie und Sprache - Biotechnologie - Technikdiskussion im Systemvergleich. Düsseldorf (VDI-Verlag) 1989

Walther Ch. Zimmerli; Volker M. Brennecke (Hrsg.): Technikverantwortung in der Unternehmenskultur. Von theoretischen Konzepten zur praktischen Umsetzung. Stuttgart (Poeschel-Verlag) 1993

Walther Ch. Zimmerli; Hansjörg Sinn (Hrsg.): Die Glaubwürdigkeit technisch-wissenschaftlicher Informationen. Düsseldorf (VDI-Verlag) 1990

Informationen der VDI-Hauptgruppe aus der Reihe "VDI-Report"

VDI-Report 3
"Der Anstellungsvertrag für Ingenieure"
2. völlig überarbeitete Auflage 1992 — DM 15.00

VDI-Report 7
"Rechtsstellung, Haftung und Verantwortung des Sicherheitsingenieurs"
VDI-Information 1978 — DM 9.50

VDI-Report 8
"Einkommen der Ingenieure in Deutschland"
VDI-Analyse 1980, mit Ergänzungsteil
"Einkommensentwicklungen 1979 - 1982" — DM 12.00

VDI-Report 10
"Einkommen der Ingenieure in Deutschland"
VDI-Analyse 1986 — DM 11.00

VDI-Report 11
"Von den zehn Geboten zu Verhaltenskodizes für Manager und Ingenieure" — DM 15.00

VDI-Report 12
"Einkommen der Ingenieure in Deutschland 1988" — DM 20.00

VDI-Report 13
"Einkommen der Ingenieure in Deutschland 1989" — DM 25.00

VDI-Report 14
"Einkommensanalyse 1991, Ingenieurgehälter in Deutschland" — DM 25.00

VDI-Report 15
"Technikbewertung - Begriffe und Grundlagen.
Erläuterungen und Hinweise zur VDI-Richtlinie 3780" — DM 18.00

VDI-Report 16
"Einkommensanalyse 1992, Ingenieur-Gehälter in Deutschland" — DM 30.00

VDI-Report 17
"Künstliche Intelligenz - Leitvorstellungen und Verantwortbarkeit" — DM 18.00

VDI-Report 18
"Einkommensanalyse 1993, Ingenieurgehälter in Deutschland" — DM 25.00

VDI-Report 19
"Unsere Verantwortung für eine umweltverträgliche Technikgestaltung.
Von abstrakten Leitsätzen zu konkreten Leitbildern" — DM 20.00

VDI-Report 20
"Hochschulbildung und Ingenieurberuf" — DM 20.00