

KÜNSTLICHE INTELLIGENZ

LEITVORSTELLUNGEN UND
VERANTWORTBARKEIT

BAND 2: TAGUNGSBERICHT

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
Dr.-Ing. Jürgen Seetzen
Prof. Dr. Ipke Wachsmuth

Fachliche Erarbeitung:

VDI-Ausschuß "Künstliche Intelligenz"

**Koordination der
Tagung:**

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.--)

© VEREIN DEUTSCHER INGENIEURE VDI 1994

Alle Rechte, auch das des auszugsweisen Nachdrucks, der auszugsweisen fotomechanischen Wiedergabe (Fotokopie) und das der Übersetzung vorbehalten.

Inhalt gedruckt auf chlorfrei-gebleichtem umweltfreundlichem Papier.

KÜNSTLICHE INTELLIGENZ

LEITVORSTELLUNGEN UND
VERANTWORTBARKEIT

BAND 2: TAGUNGSBERICHT

Inhalt

Leitidee und Ziel der Tagung.....	1
<i>Prof. Dr. A.B. Cremers, Universität Bonn, Leiter des Instituts für Informatik III, Obmann des VDI-Ausschusses "Künstliche Intelligenz"</i>	
Vorwort der Herausgeber.....	2
Durchsicht und Zusammenfassung der Beiträge im VDI-Report 17 (Hrsg. A.B. Cremers, R. Haberbeck, J. Seetzen, I. Wachsmuth).....	9
<i>G. Grube, Berlin</i>	
Sind die Leitvorstellungen der KI verantwortbar?.....	27
<i>Dr.-Ing. J. Seetzen, VDI/VDE-Technologiezentrum Informationstechnik, Teltow G. Grube, Berlin</i>	
Müssen die bisherigen Positionen der KI überdacht werden?.....	34
<i>Prof. Dr. J. Siekmann, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken</i>	
Round-table-Gespräch über die "harte" KI-These: Der Mensch als Informationsverarbeitungssystem?.....	39
<i>Positionsbestimmungen von Mitgliedern des VDI-Ausschusses "Künstliche Intelligenz"</i>	
<i>- Prof. Dr.-Ing. R. Eckmiller, Universität Düsseldorf, Institut für Physikalische Biologie</i>	
<i>- Prof. Dr. G. Görz, Universität Erlangen-Nürnberg, IMMD</i>	
<i>- Prof. Dr. S. Krämer, Institut für Philosophie der FU Berlin</i>	
<i>- Prof. Dr. G. Strube, Institut für Informatik und Gesellschaft der Universität Freiburg</i>	
<i>Moderator</i>	
<i>Prof. Dr. I. Wachsmuth, Universität Bielefeld, Technische Fakultät</i>	

Künstliche Intelligenz als Technik - Hoffnungen, Erwartungen und heutiger Stand.....	52
<i>Prof. Dr. G. Barth, Daimler-Benz AG, Ressort Forschung und Technik, Ulm</i>	
Probleme und Entwicklungen der Neuroinformatik.....	59
<i>Prof. Dr. W. von Seelen, Ruhr-Universität Bochum, Institut für Neuroinformatik</i>	
Möglichkeiten und Grenzen der KI im Ingenieurbereich - Überlegungen aus der Sicht der Automatisierungstechnik.....	70
<i>Prof. Dr.-Ing. J. Lunze, TU Hamburg-Harburg, Arbeitsbereich Regelungstechnik</i>	
Leitbilder der Technik - ihre Bedeutungen, Funktionen und Potentiale für den KI-Diskurs.....	83
<i>Prof. Dr. M. Dierkes, Wissenschaftszentrum Berlin für Sozialforschung</i>	
Zukunftsvisionen der KI: Wünschbare und realistische Wirkungen auf die Lebenswelt.....	99
<i>Prof. Dr. I. Wachsmuth, Universität Bielefeld, Technische Fakultät</i>	
Zukunftsvisionen der KI: Wünschbare und realistische Wirkungen auf die Arbeitswelt.....	106
<i>Prof. Dr. W. Coy, Universität Bremen, Fachbereich Mathematik/ Informatik</i>	
Round-table-Gespräch: Konsequenzen für eine industrielle Technologiefolgenabschätzung der KI.....	109
<i>Statements von</i>	
- <i>R. Haberbeck M.A., Siemens Nixdorf Informationssysteme, Berlin</i>	
- <i>Dr. R.A. Müller, Daimler-Benz AG, Ressort Forschung und Technik, Berlin</i>	
- <i>Prof. Dr. O. Herzog, Universität Bremen, Fachbereich Informatik</i>	
- <i>Dr. G. Klett, BASF AG, Abt. Zentrale Informatik-Technologie, Ludwigshafen</i>	
<i>Moderator</i>	
<i>Dr. R. Stransfeld, VDI/VDE-Technologiezentrum Informationstechnik, Teltow</i>	
Anhang: Mitglieder des VDI-Ausschusses "Künstliche Intelligenz" und Gastreferenten.....	118

**VDI-Symposium
"Künstliche Intelligenz - Leitvorstellungen und Verantwortung",
Gustav-Stresemann-Institut, Bonn-Bad Godesberg,
10. - 11. September 1992**

Leitidee und Ziel der Tagung

Teilbereiche des Forschungsfeldes "Künstliche Intelligenz" (KI) haben bereits in vielen technischen Anwendungsgebieten praktische Bedeutung erlangt (wissenbasierte Software-technik, Robotik, Mustererkennung), von weiteren Teilgebieten (Sprachverarbeitung, intelligente Kommunikationstechnik) wird dies in den nächsten Jahrzehnten erwartet. Heute sind nicht nur Fragen nach der gesellschaftlichen Akzeptanz einer entwickelten KI zu stellen, sondern vorerst auch die grundsätzlichen Perspektiven und technischen Realisierungsmöglichkeiten dieses Forschungsgebietes zu diskutieren.

Da die KI das Ziel verfolgt, die kognitive Intelligenz von Menschen maschinell simulieren zu können, ist sie in besonderer Weise auf die anthropologischen Disziplinen (Philosophie, Kognitionspsychologie, Linguistik, Neurophysiologie etc.) angewiesen. In diesem wechselseitigen wissenschaftlichen Forschungs- und Erkenntnisprozeß besteht die Chance einer Technikgestaltung, die sich später als sozialverträglich erweisen sollte.

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 der ingenieurmäßigen Umsetzung der KI-Forschung bereits jetzt genauer abschätzen.

Prof. Dr. Armin B. Cremers,
Obmann des VDI-Ausschusses "Künstliche Intelligenz"

Vorwort der Herausgeber

Die Arbeit des VDI-Ausschusses "Künstliche Intelligenz" ist ein Beitrag zum Diskurs über die Technikfolgenabschätzung und Technikbewertung der Verfahren zur Künstlichen Intelligenz (KI).

Der Ausschuß beschäftigt sich seit seiner Gründung im Dezember 1988 mit grundlegenden Fragen nach Perspektiven und Auswirkungen der Modellierung und Simulation von Fähigkeiten der natürlichen Intelligenz durch informationsverarbeitende Maschinen.

KI-Forschung kann nur interdisziplinär durchgeführt werden: Durch die Zusammensetzung des VDI-Ausschusses durch ausgewiesene Spezialisten der einzelnen Sachgebiete (Ingenieurwissenschaften, Informatik, Erkenntnis- und Wissenschaftstheorie, Kognitionspsychologie, Philosophie, Linguistik, Neurowissenschaften etc.) ist es möglich, ein Themenspektrum unterschiedlicher Disziplinen zusammenzufassen und zu diskutieren.

Bei dem gestellten Thema "Das Menschenbild in der KI" geht es um die vielfach vorzufindende Annahme einer grundsätzlichen Vergleichbarkeit der Informationsverarbeitung von Mensch und Maschine. Der Ausschuß diskutiert die expliziten wie auch die vage vorhandenen Vorstellungen der KI-Forschung über den Menschen, um damit Aufschluß über die grundsätzliche Richtung dieses Forschungszweiges zu gewinnen.

Der Ausschuß "Künstliche Intelligenz" ist aus dem Ausschuß "Sprache und Informationstechnologie" hervorgegangen, der unter der Leitung von W. Strombach stand, und schließt an dessen Arbeiten an. Sowohl von der inhaltlichen Entwicklung her als auch hinsichtlich der nun ausgeweiteten Interdisziplinarität handelt es sich um eine genuine Aufgabenstellung für den VDI.

Der Bundesminister für Forschung und Technologie (BMFT) hat 1989 die technisch-wissenschaftlichen Vereinigungen ermutigt und ermuntert, die dringend erforderliche Diskussion um die Möglichkeiten und Folgen der in der Entwicklung befindlichen Verfahren der künstlichen Intelligenz, besonders der Expertensysteme, voranzubringen, um bei dieser disziplinübergrei-

fenden technischen Entwicklung den öffentlichen, rationalen Diskurs rechtzeitig aufzunehmen. Dem Ausschuß ist zur Unterstützung seiner Arbeit ein Förderprojekt vom BMFT bewilligt worden.

Ausgangslage

Künstliche-Intelligenz-Forschung ermöglicht es, 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 (KI) vor allem ein Bild seiner Intelligenz, wobei die Intelligenzauffassung entscheidend von der Sichtweise eines informationsverarbeitenden Systems geprägt ist.

Künstliche-Intelligenz-Forschung zielt auch darauf ab, bei der Konstruktion informationsverarbeitender Systeme kognitive Funktionen einzubeziehen, um die theoretisch entwickelten Konzepte und Techniken nutzbringend umzusetzen. Auf informationsverarbeitende Maschinen werden Eigenschaften menschlicher Intelligenz, etwa Schlußfolgerungsfähigkeiten, übertragen, um dadurch geistige Tätigkeiten des Menschen zu unterstützen, zu verstärken oder teilweise zu ersetzen.

Im Rahmen der Sprachverarbeitung geht man von der symbolischen Darstellung und Verarbeitung von Sprache aus. Erst neuerdings werden Konzepte neuronaler Netze stärker in die Computerlinguistik einbezogen. Bei den Deduktionsverfahren setzt man voraus, daß Wissensverarbeitung grundsätzlich formalen logischen Verfahren folgt, die die symbolischen Sprachdarstellungen verknüpft. Im Rahmen der Expertensystemtechnik versucht man diese Wissensverarbeitung praktisch zu nutzen. Insgesamt bezieht sich die Künstliche Intelligenz auf den kognitiven Bereich der "Intelligenz", was aber für reale menschliche Intelligenz nicht erschöpfend ist.

Zielsetzung

Wie bei jeder Technologie ergibt sich die Frage nach der Betroffenheit, wobei durch den Anspruch, geistige Tätigkeiten des Menschen zu formalisieren, eine neue Qualität der Folgenabschätzung technischer Mittel berührt wird.

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

1. *Wo steht der anthropologisch-philosophische 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 deutsche Philosophie nach der Aufklärung und dem Idealismus (Kant, Hegel) seit Mitte des vorigen Jahrhunderts durch Schopenhauer, Nietzsche, Husserl und in diesem Jahrhundert durch Wittgenstein, Heidegger, Jonas, Gadamer sowie in Frankreich durch Foucault, Derrida und Lyotard herausgebildet hat, mit den meist implizit vertretenen rationalistisch-positivistischen Grundvorstellungen vieler KI-Entwickler über den Menschen in Übereinstimmung oder im Gegensatz steht. Es zeigt sich, daß es gravierende Unterschiede in den Sichtweisen gibt, die sich, sei es in der Sprachphilosophie, sei es 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 aber 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 modernen Anthropologie und Philosophie einschließlich der Wissenschaftstheorie und der Linguistik den Grundannahmen der Kognitionswissenschaften 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, daß prinzipielle Grenzen dieser Entwicklungen rechtzeitig in den Blick kommen.

2. *Welche "Menschenbilder" spiegeln sich in den Voraussetzungen der Bemühungen um die Entwicklung der Künstlichen Intelligenz?*

Es könnte sich herausstellen, daß die Wunschvorstellungen vieler KI-Protagonisten hinsichtlich der Möglichkeiten der KI unrealistisch 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 und Technikwissenschaftlern erforderlich, um die notwendige Orientierung zu erarbeiten.

Betrachtet man die gegenwärtige KI-Forschung 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 nur diffuse Hinweise auf ein spezifisches Menschenbild, es sei denn, man betrachtet bereits den überstrapazierten Vergleich des menschlichen Gehirns mit einem Computer als eine entsprechend aussagekräftige Metapher.

In dem klassischen Ansatz der KI, nämlich dem der Symbolverarbeitung, 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 jüngerer Zeit entstandenen Ansätze zur Simulation neuronaler Strukturen, die beispielsweise unter dem Etikett "Konnektionismus" bekannt geworden sind, stellen demgegenüber einzig einen methodologischen, jedoch keinen erkenntnistheoretischen Unterschied dar, wird doch auch hier von einer automatentheoretischen Modellierungsprämisse ausgegangen.

Die Simulation von Verstandesleistungen auf einer Maschine setzt bestimmte Grundannahmen über die jeweiligen mentalen Vorgänge beim Menschen voraus, die zum Teil auf der individuellen Introspektion der involvierten KI-Forscher beruhen, zum Teil der Analyse entsprechender philosophischer, psychologischer, linguistischer oder neurobiologischer Literatur und Forschung entstammen. Die genaue Betrachtung der sich hier findenden Vorstellungen der einzelnen Bereiche kann ein Menschenbild verdeutlichen, dessen erstes Charakteristikum bereits in der angenommenen Möglichkeit einer Zersplitterung kognitiver Prozesse besteht. Aus dieser Sichtweise hat sich ein methodischer Zugang für die Bearbeitung der obengenannten Fragestellung ergeben:

- a) die Identifikation der impliziten und expliziten Annahme 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 Menschenbild der KI zu gelangen und dessen philosophische, linguistische und neurobiologische Grundlagen

zu bestimmen und damit auch einen Ansatzpunkt für die Lösung anwendungsorientierter Fragen zu gewinnen.

3. *Welche Änderungen des Menschenbildes ergeben sich unter dem Einfluß einer "entwickelten" KI?*

Solche Fragen betreffen die Gestaltung dieser neuen Technologien 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 Weltansichten informationsverarbeitende Systeme so zu gestalten, daß sie dem Menschen nicht undurchschaubar und fremd sind?

In diesem Zusammenhang sind folgende exemplarische Fragen zu stellen:

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

Zusammensetzung und Arbeitsweise des Ausschusses "Künstliche Intelligenz"

Der Verein Deutscher Ingenieure hat angesichts dieser Herausforderungen im Rahmen seiner Hauptgruppe "Der Ingenieur in Beruf und Gesellschaft" im Bereich "Mensch und Technik" den interdisziplinär besetzten Ausschuß "Künstliche Intelligenz" angesiedelt. Der Ausschuß besteht z.Zt. aus 20 Mitgliedern. Die Vertreter der ingenieur- sowie der geisteswissenschaftlichen Disziplinen stehen dabei in einem ausgewogenen Verhältnis. Die Zusammensetzung des Ausschusses ermöglicht es, zu allen relevanten Fragen der KI ausgewiesene Wissenschaftler sowie Anwender in der Industrie zu hören.

In einer Vorbereitungsphase hatte sich der Ausschuß mit folgenden Einstiegsthemen im Rahmen von Vortrags- und Diskussionsveranstaltungen befaßt:

- Computersimulation von Erkenntnisprozessen - eine Herausforderung für die Philosophie (F. Rapp, Dortmund)
- Neuroinformatik: Informationsverarbeitung in biologischen und technisch neuronalen Netzen (R. Eckmiller, Düsseldorf)
- Das Menschenbild angesichts der informationstechnischen Entwicklungen (J. Seetzen, Berlin)
- Neuroästhetik und Künstliche Intelligenz (I. Rentschler, München)

Zur Durchführung der eigentlichen Projektarbeit seit 1990 hat sich der Ausschuß in drei Arbeitsgruppen gegliedert, entsprechend den drei Leitfragen:

1. Anthropologischer Diskurs, (Leitung: J. Seetzen)
2. Menschenbilder in der gegenwärtigen Künstlichen Intelligenz, (Leitung: R. Haberbeck)
3. Gesellschaftlicher Einfluß einer entwickelten Künstlichen Intelligenz, (Leitung: I. Wachsmuth)

Neben den turnusmäßigen Ausschußsitzungen, die zwei- bis dreimal im Jahr stattgefunden haben, sind zwei ausschußinterne Workshops veranstaltet worden, um die Diskussion zwischen den Mitgliedern zu intensivieren, und zwar am 2.-3.11.1990 in Berlin und am 6.-8.5.1991 in Bad Homburg. Zur Vorbereitung dieser Workshops sind von den Gruppen Arbeitspapiere erstellt worden, die zum einen die einschlägige Literatur im Hinblick auf die Fragestellung des Ausschusses aufbereitet, zum anderen einen Aufriß der Debatte geleistet haben. Diese Papiere sind unter Einbeziehung externer Mitarbeiter über Werkverträge vorbereitet worden (M. Butt, G. Grube, G. Lischka, M. Wilker). Die auf den Workshops entwickelten Diskurse sind sorgfältig protokolliert worden, um sie der eigenen Auswertung oder auch späteren Analysen zugänglich zu machen.

Als vorläufiger Zielpunkt des Projektes ist das Symposium über Leitvorstellungen und Verantwortbarkeit der Künstlichen Intelligenz in Bad Godesberg vom 10.-11.9.1992 geplant worden. Das Hauptanliegen besteht darin, die fachliche und sonstige Öffentlichkeit über die Ergebnisse der Ausschußarbeit zu informieren und die stattfindende Diskussion explizit in den Projektdiskurs einzubeziehen. Zu diesem Zweck ist in Form des VDI-Reports No. 17 eine vorläufige Zusammenfassung der Projektresultate herausgegeben worden. Mit der Durchsicht

und Zusammenfassung der Beiträge im VDI-Report 17 ist G. Grube beauftragt worden, dessen Aufsatz sich hieran anschließt.

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

Durchsicht und Zusammenfassung der Beiträge im VDI Report 17

(Hrsg. A.B. Cremers, R. Haberbeck, J. Seetzen, I. Wachsmuth)

G. Grube

1. Report 17 und Diskurs

Dem vorliegenden Tagungsbericht ist die Veröffentlichung des VDI Reports 17 „**Künstliche Intelligenz. Leitvorstellungen und Verantwortbarkeit**“ vorausgegangen. Ebenso wie die Tagung in Bonn, die dem Tagungsbericht zugrunde liegt, ist der VDI Report 17 aus dem Diskursprojekt „Das Menschenbild in der Künstlichen Intelligenz“ hervorgegangen. Neben dem Report 17 gibt es eine Reihe von Diskursprotokollen, die den Verlauf des gesamten Diskursprojektes dokumentieren. Die nachfolgende Durchsicht und Zusammenfassung bezieht sich nicht auf Diskursprotokolle, sondern nur auf die Beiträge im Report 17. Da diese Beiträge aber im Kontext des Diskursprojektes entstanden sind, möchte ich vorab einige Gesichtspunkte hervorheben.

Erstens: Die Protokolle zeigen die Notwendigkeit eines interdisziplinären Diskurses sehr deutlich, da sie zum einen die Differenzen in den Auffassungen der Wissenschaftler dokumentieren und zum anderen auch das Gewicht dieser Differenzen beleuchten. So heißt es zum Beispiel im Resümee des Diskursprotokolls II-2 zu den verschiedenen Auffassungen zum Informationsbegriff: „Dabei handelt es sich keineswegs um leicht zu überwindende Mißverständnisse, oder um einfach anzuleichende Positionen; vielmehr zeigte sich, daß die Unterschiedlichkeit der Meinungen in tieferen philosophischen Positionsunterschieden gründet.“ (II-2, S. 30) Die Beiträge im Report 17 bestätigen diese Notwendigkeit, eine Auseinandersetzung zwischen tief angelegten und scheinbar fest verankerten Positionsunterschieden zu führen.

Zweitens: Im Diskursprotokoll II-1 wird festgestellt, daß es in der Künstlichen-Intelligenz-Forschung gegenwärtig kaum „explizite Vorstellungen über den Menschen“ gibt, die den Forschungsbestrebungen zugrunde liegen, „so finden sich in der Regel allemal diffuse Hinweise auf ein Menschenbild“ (II-1, S.4), und es wird zum Ziel des Diskurses erklärt, die expliziten und vagen Vorstellungen zum Menschenbild der Künstlichen Intelligenz (KI)

herauszuarbeiten, um „Hinweise auf spätere Gestaltungsperspektiven sowie mögliche Auswirkungen zu geben“ (II-1, S.3). Die Dokumentation der verschiedenen Auffassungen und Diskurspositionen durch den Report 17 kann als ein Beitrag zur Intention des interdisziplinären Diskurses angesehen werden, da allein durch die Gegenüberstellung expliziter und vager Vorstellungen zum Menschenbild Folgerungen gewonnen werden können, die die „große“ Diskussion (die gesellschaftliche, wissenschaftliche und forschungspolitische) mit erhellen. Dabei bringen Gegenüberstellungen der Beiträge des Reports Kontroversen ans Licht, denen auch schon in den Diskursprotokollen ein wesentlicher Stellenwert zugebilligt wird. Beispielsweise kann man anhand der Beiträge im Report einer wesentlichen Fragestellung nachgehen, die im Protokoll II-1 wie folgt formuliert ist: „Es könnte sich herausstellen, daß die Wunschvorstellungen der KI-Protagonisten hinsichtlich der Möglichkeiten der KI unreal sind oder daß das Menschenbild, das sich im anthropologisch-philosophischen Diskurs gebildet hat, angesichts der KI-Entwicklung der Revision bedarf.“ (II-1, S.4)

Drittens: Thematische Entwicklungen, die in den Protokollen niedergelegt sind, spiegeln sich im Report wider. So gewinnen etwa die Fragen nach „Verantwortung“ im Verlauf des Diskurses an Gewicht. Dabei sind diese Fragen im philosophischen Diskurs mit den Begriffen der Personalität und der Subjektivität verknüpft, vor dem Hintergrund der „real existierenden“ KI hingegen mit der Debatte um die qualitative Trennung zwischen KI und konventioneller Informationstechnik verbunden. Im Diskurs zu den „Zukunftsauswirkungen der Künstlichen Intelligenz“ wurde die Frage diskutiert, „ob Maschinen in absehbarer Zeit Personenstatus zugerechnet werden kann bzw. ob Maschinen als Subjekte begriffen werden können.“ (II-2, S.31)

Viertens: In einigen Punkten, so dokumentieren die Protokolle, war im Diskurs Konsens zu erzielen. Das Interessante ist, daß die Beiträge im Report 17 zum Teil eine Struktur aufweisen, über die im Diskurs ein Konsens erreicht wurde. Hierbei ist besonders hervorzuheben, daß es in einer KI-Debatte, gerade wenn es sich um eine interdisziplinäre handelt, darauf ankommt, zu klären, ob es um die technischen Leistungen oder um philosophische Ansprüche der KI geht. Im Protokoll II-1 heißt es: „allgemein wird anerkannt, daß ein technisches von einem theoretischen Interesse zu unterscheiden ist.“ (II-1, S.39)

2. Zur Durchsicht

Mit der Durchsicht und Zusammenfassung der Beiträge im VDI Report 17 wird versucht, vor allem die Positionen herauszulesen, die in den Beiträgen jeweils zum Menschenbild der KI eingenommen werden.

Die Beiträge im **Report 17** sind sehr heterogen. Das Spektrum reicht von kurzen philosophischen Einführungen (Kemmerling zum Begriff des Funktionalismus) über sehr pointierte Stellungnahmen (Eckmiller über neue Welt- und Menschenbilder) bis hin zu gründlichen Spekulationsversuchen (Görz et al. über die Folgen für eine Gesellschaft mit hoch entwickelter KI). Die große Unterschiedlichkeit der Beiträge spiegelt sich aber nicht in der inhaltlichen Gliederung des Reports wider. D.h. die drei Abschnitte (**I: Menschenbild und Computeranthropologischer Diskurs**, **II: Menschenbilder in der real existierenden KI**, **III: Zukunftsauswirkungen der Künstlichen Intelligenz**), die ja für das Diskursprojekt, in dessen Rahmen der Report entstanden ist, auch pragmatische Relevanz hatten, entsprechen nicht unbedingt den konzeptionellen Voraussetzungen, die ein Autor seinem Beitrag, vielleicht auch der Autor als Diskursteilnehmer seiner Auseinandersetzung, zugrunde gelegt hat. So würde man etwa den Beitrag über die Neuroinformatik von Eckmiller in Abschnitt II, da er auch sehr spekulativ ist, lieber in Abschnitt III unterbringen. Oder man möchte für viele Beiträge, die im Report über alle Abschnitte verteilt sind, sich aber überproportional mit der Verantwortungsproblematik befassen, vielleicht einen eigenen Abschnitt einrichten.

Tatsächlich lassen sich aber konzeptionelle Voraussetzungen finden, die die Beiträge in einen systematischen Zusammenhang bringen. Diese Systematik stimmt jedoch nicht mit den Abschnitten des Reports überein, sondern ist nach drei Aspekten ausgerichtet, die im Gesamtkontext aller 16 Beiträge eine herausragende Rolle spielen. Diese Aspekte sind a) das Menschenbild der KI, b) die Verantwortung und c) die technischen Leistungen der KI.

Mit Hilfe der erwähnten Systematik lassen sich die 16 Beiträge in zwei Hauptgruppen teilen. Zur einen Hauptgruppe gehören drei Beiträge (Capurro/Seetzen, Krämer, Seetzen/Capurro), die die KI als Ausdruck eines Menschenbildes thematisieren, ohne die möglichen technischen Leistungen zu berücksichtigen. Zur anderen Hauptgruppe gehören diejenigen Beiträge (alle übrigen), die in erster Linie die KI-Technik und nur in zweiter Linie ein zugehöriges Menschenbild thematisieren. Diese große Hauptgruppe läßt sich in drei Untergruppen aufteilen.

Die eine Untergruppe könnte als die visionäre bezeichnet werden (Eckmiller, Görz et al., Schreiber). Es gehört zur visionären Position, daß das „alte“ Menschenbild¹ mit Hilfe verbesserter KI-Technik von gegenwärtig entdeckten Mängeln befreit wird (besonders Eckmiller, Schreiber) und die Bedeutung des Menschen in dem neu gekleideten „alten“ Menschenbild stark relativiert ist (besonders Eckmiller, Görz et al.). Die nächste Untergruppe wird von den Beiträgen gebildet, die den technischen Möglichkeiten der KI keine Grenzen stecken, mit anderen Worten, die visionäre Position nicht ausschließen, aber gerade wegen der unabsehbaren

¹ In der nachfolgenden Zusammenfassung zeigt sich, wie in einzelnen Beiträgen das „alte“ Menschenbild der KI verstanden wird - siehe z.B. Capurro/Seetzen und Müller.

technischen Möglichkeiten den Verantwortungsbegriff formulieren (Strube, Haberbeck, Kremer, Röpke). Die dritte Untergruppe umfaßt zunächst diejenigen Beiträge, die von den technischen Schwierigkeiten beziehungsweise Grenzen der KI ausgehen (Kemmerling, Becker/Lischka, Cremers/Eder/Hinze, Schlachetzki, Wender, Müller). Sie ist aber hinsichtlich der Reaktion noch einmal in zwei Lager aufzugliedern. Das eine Lager bilden die Beiträge (Becker/Lischka, Müller), in denen für ein anderes Menschenbild argumentiert wird, das auch eine verbesserte Technik nach sich ziehen sollte. Während im anderen Lager (Kemmerling, Cremers/Eder/Hinze, Schlachetzki, Wender) aufgrund der technischen Schwierigkeiten der KI lediglich eine Skepsis gegen das „alte“ Menschenbild artikuliert wird.

In dieser Systematik zerfällt die erstgenannte Hauptgruppe (Capurro/Seetzen, Krämer, Seetzen/Capurro) in keine Untergruppen, da es hier jedesmal um KI geht, durch die ein Menschenbild (das „alte“) gestützt wird, das moralische Dimensionen beziehungsweise das Verantwortungsvermögen eines Menschen vernachlässigt. Die nachstehende Abbildung zeigt die Systematik.

1 Menschenbild in der KI, Verantwortung	Capurro/Seetzen, Krämer, Seetzen/Capurro			
2 technische Leistungen der KI	2.1 Vision und Risiko, Ver- antwortung	Strube, Haberbeck, Kremer, Röpke		
	2.2 Vision	2.2.1 Aufwertung des "alten" Menschenbildes	Eckmiller, Schreiber	
		2.2.2 Relativierung des Men- schen in „neuem“ Men- schenbild	Eckmiller, Görz et al.	
	2.3 Grenzen der KI-Technik	2.3.1 mit „neuem“ Menschen- bild Überwindung der Grenzen	Becker/Lischka, Müller	
		2.3.2 Skepsis gegen das „alte“ Menschenbild der KI	Kemmerling, Cremers/Eder/Hinze, Schlachetzki, Wender	

Abbildung 1: Die Systematik

Hinter der Zusammenfassung, die alle 16 Beiträge berücksichtigt, steht die Absicht, jeden der Beiträge derart auf das Wesentliche zu reduzieren, daß sich die Beiträge mehr oder weniger alle aufeinander beziehen lassen.

Im folgenden beziehen sich die Angaben in Klammern, (A2) z.B., auf den Absatz des zusammengefaßten Beitrags. Bei manchem Zitat, oder wenn es sich um einen Absatz eines anderen Beitrags handelt, folgt auf die Absatznummer eine Seitenzahl.

Bei einem Beitrag, der von mehreren Autoren verfaßt ist, heißt es der Übersichtlichkeit halber (wie bereits weiter oben) z.B. Becker/Lischka für Becker und Lischka oder Görz et al. für Görz, Kremeier, Röpke, Schreiber, Strube, Wachsmuth und Wilker.

3. Zusammenfassung

Capurro/Seetzen argumentieren in ihrem Beitrag² „**Menschenbilder**“, daß die Anstrengungen - besonders die philosophischen -, ein Menschenbild zu bestimmen, seit Kant deutlich zwei Grundfragen folgten, nämlich zum einen der epistemischen Frage: „Was kann der Mensch wissen?“ und zum anderen der ethischen Frage: „Was soll der Mensch tun?“ (A3). Mit dem Auftreten der modernen Naturwissenschaften sei die ethische Frage in den Schatten der epistemischen geraten (A4). KI sei nun eine Technik, deren Anwendungsbereich letztlich menschliches Wissen ist und die in den Grenzen der naturwissenschaftlichen Tradition entwickelt wird (A6). So trage die KI-Entwicklung zur Festigung eines einseitig naturwissenschaftlichen Menschenbildes bei. Die Kritik von Capurro/Seetzen zielt auf die Einseitigkeit des Menschenbildes ab und fordert dessen Ergänzung um die ethische Fragestellung. Auch stützen sie ihre Kritik auf einen Informationsbegriff, der die KI, sofern es sich dabei um eine Informationstechnik handelt, mit humanen Eigenschaften in Verbindung bringt, die eine große Rolle bei ethischen Fragen spielen. Denn Information sei eine auf Physik nicht reduzierbare „Kategorie der lebenden Strukturen und Systeme“ (A7), von welcher unser Entscheidungsvermögen und verantwortliches Handeln abhängen.

In ihrem Beitrag „**Ethischer Ausblick**“³ markieren **Seetzen/Capurro** einige Hauptpunkte ihrer Kritik an dem in ethischer Hinsicht defizitären Menschenbild der KI. Die Naturwissenschaftler und Techniker übersähen, daß ethische Belange nicht auf Fragen des Wissens reduziert werden können, sondern diese gerade auch von Fragen des Wollens und der Emotionen abhängen (A5). Darüberhinaus trage die Informationstechnik modernen Stils dazu bei, daß Ethik nicht mehr nur eine Frage individuellen Verhaltens, sondern der kollektiven Verantwortung ist. Kollektive Verantwortung liege bei den „Organisationen, Unternehmen, Verwaltungen, Parteien, Regierungen“ usw.

²S.11-S.14, 11 Abs.

³S.49-S.51, 9 Abs.

KI wird von Capurro/Seetzen nicht weiter differenziert. Ihre Einschätzung wird durch den Beitrag⁴ „**Künstliche Intelligenz und ihre technisch-physikalische Realisierung**“ von **Schlachetzki** bestätigt, in dem er darlegt, daß allen KI-Ansätzen, gleichgültig ob es bei ihnen um digitale, analoge, optische Rechner, um Fuzzy Logic oder Neuronale Netze geht (A8-A13), ein mechanistisches Weltbild zugrunde liegt. Jede realisierte KI-Technik sei in dem Sinne deterministisch, in dem auch alle Aspekte, die ein mechanistisches Weltbild umfaßt, physikalisch determiniert seien (A20). Dieses Weltbild schließe ein mechanistisches Menschenbild ein (A21). Geht man davon aus, daß Schlachetzkis mechanistisches Menschenbild und das naturwissenschaftliche Menschenbild von Capurro/Seetzen vergleichbar sind, so stimmen hier die Auffassungen zum Menschenbild der KI überein. Erst mit der Bewertung dieser Auffassung endet die Übereinstimmung. Denn Schlachetzki hält es aufgrund mangelnder Beweise für eine Ansichtssache, ob das mechanistische das richtige Menschenbild ist (A22), obwohl er die Intelligenz im Rechner als Verarbeitung von Information im technischen Sinne bezeichnet (A21) und auch einräumt, daß mit dem Rechner Phänomene wie Bewußtsein, Gefühl und Wille nicht realisiert werden können (A23). Dementsprechend dürfe Rechnern zwar nicht die Kompetenz abgesprochen werden, Entscheidungen zu treffen, sie ihnen aber auch nicht ohne weiteres überlassen werden, da die Testverfahren zu unzulänglich seien (A24).

Der Beitrag⁵ „**Die Rolle der mathematischen Logik in der Künstlichen Intelligenz**“ von **Cremers/Eder/Hinze**, der nichts explizit über ein Menschenbild in der KI aussagt, bestätigt lediglich das mechanistische beziehungsweise naturwissenschaftliche Menschenbild der KI, sofern den Prinzipien der klassischen Logik entsprochen wird. Und obwohl Cremers/Eder/Hinze einräumen, daß KI-Systeme mittels ihrer Informationsverarbeitungsmechanismen neues Wissen zu erschließen vermögen (A4), weisen sie auf Grenzen der mathematischen Logik hin, die gegebenenfalls als ungerechtfertigte Grenzen eines Menschenbildes interpretiert werden können, dem die KI über die Prinzipien der klassischen Logik verpflichtet ist. Zum einen liege die Begrenzung in der Voraussetzung, daß mathematische Logik nur dort angewandt werden könne, wo der Anwendungsbereich mathematisch formalisierbar sei, und zum anderen in der Voraussetzung, daß Beweise in zu großen Suchräumen ohne die bewährte Hilfe der Intuition gefunden werden müßten (A6). Die Reichweite eines Menschenbildes, das den Prinzipien der mathematischen Logik entspricht, das dem bislang thematisierten KI-Menschenbild nicht widerspricht, hört bei dem Phänomen Intuition auf. Obwohl sich die Intuition der mathematischen Formalisierbarkeit entzieht, enthalten sich Cremers/Eder/Hinze einer Bewertung. Sie schreiben zum einen, daß es Hinweise gebe, „daß co-NP-vollständige Pro-

⁴S.72-S.82, 25 Abs.

⁵S.60-S.65, 12 Abs.

bleme“ generell nicht automatisch lösbar seien, räumen aber zum anderen ein, daß die „Anwendung des automatischen logischen Schließens höchstwahrscheinlich“ begrenzt bleibe, „solange“ die Intuition vom Computer nicht nachvollzogen werden könne (A12).

Gegenüber Capurro/Seetzen, Schlachetzki und Cremers/Eder/Hinze ist es **Eckmiller** in seinem Beitrag⁶ „**Neuroinformatik und Künstliche Intelligenz**“ wichtig, KI ihren Ansätzen nach zu unterscheiden. Er räumt ein, daß sich die Ansätze in ihrer Zielsetzung, nämlich die "Informationsverarbeitungs-Leistungen" von Menschen und Tieren auf technische Systeme zu übertragen, gleichen (A1). Der Unterschied bestehe darin, daß der eine Ansatz eine funktionelle Perspektive einnimmt (A2), während mit dem von ihm vertretenen Ansatz der Neuroinformatik versucht wird, die realen Hirnfunktionen in technische Systeme zu übertragen (A3). Gründe, die für den neuroinformatischen Ansatz sprächen, sind das Versagen des funktionellen Ansatzes (A5), die Fortschritte in der Hirnforschung (A6) und die überragende Tatsache, daß sich das zentrale Nervensystem mit Erfolg bewährt (A11). Trotz Eckmillers Differenzierung der KI-Techniken scheint das Menschenbild, das er der Neuroinformatik zuschreibt, nicht von dem abzuweichen, welches Capurro/Seetzen und Schlachetzki für die KI vorschlagen haben. Eckmiller fügt den Beschreibungen dieses Menschenbildes den Vergleich des Menschen mit einem „Automobil“ hinzu (A12). Als Perspektive für den Menschen sieht er gleichzeitig durch die Neuroinformatik die Chance zu neuen Menschenbildern eröffnet (A14). In diesem Sinne schreibt er: „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.“ (A18, S.70) Das von Eckmiller durchgespielte Menschenbild weicht tatsächlich in einer Hinsicht von dem „alten KI“-Menschenbild ab, denn er verallgemeinert es und verbindet mit seinem neuen „alten“ Menschenbild, daß es auch ein angemesseneres Menschenbild für den Menschen (und nicht lediglich den Computer beziehungsweise den Menschen als Computer) ist.

Damit zeigt sich bereits eine Polarisierung der Positionen. Das Menschenbild, das Capurro/Seetzen der KI als einseitig epistemisch unterstellen, wird von Eckmiller als Menschenbild schlechthin projiziert, das allerdings von der 'funktionellen' KI technisch nicht umgesetzt werden konnte. Schlachetzki und Cremers/Eder/Hinze beziehen explizit keine Position, sondern sehen die Schwierigkeiten, aus mechanistischer beziehungsweise der mathematisch logischen Perspektive Phänomene wie Wille oder Intuition zu erklären. Capurro/Seetzen beziehen aus der Existenz dieser Phänomene ihre Kritik an dem einseitigen Menschenbild, während Eckmiller gerade die Hoffnung hat, daß derartige kritische Potentiale

⁶S.66-S.71, 18 Abs.

hinsichtlich des in Rede stehenden Menschenbildes bald keine gewichtigere Bedeutung mehr haben werden.

Kemmerling macht in seinem Beitrag⁷ „**Der Geist als Computer**“ eine Bemerkung, „Geist und Hirn sind etwa so verschieden wie Blick und Auge“ (A28, S.25), mit der er der Position Eckmüllers direkt widerspricht. Dabei geht Kemmerling davon aus, daß das Hirn neurobiologisch, d.h. physikalisch zu erklären sei, während geistige Phänomene nicht auf Physik zu reduzieren seien. Kemmerling referiert das Konzept des Funktionalismus, das versucht, einem „Doppelcharakter“ der geistigen Phänomene gerecht zu werden. Seinen Doppelcharakter besitze das geistige Phänomen, da es einerseits eine irreduzible funktionale Rolle sei und andererseits als funktionale Rolle auf irgendeine Art materiell realisiert werde (A10). Die funktionalistische Position biete sich dann als Position für die KI an, wenn ein Computer als materiell realisiertes System betrachtet werde, in dem funktionale Rollen programmiert seien (A20). Hierzu aber erklärt Kemmerling, daß aufgrund des Funktionalismus KI-Leistungen höchstens als Analogie einiger geistiger Leistungen verstanden werden können, während andere geistige Phänomene wie Emotion, Wille oder gar ein Geschmackserlebnis nicht durch einen Computer zu materialisieren seien (A29, A30). Damit kann die Beziehung KI und Menschenbild um einen weiteren Aspekt ergänzt werden. Mit dem funktionalistischen Ansatz, der mit einem Menschenbild verträglich ist, in dem geistige Phänomene nicht reduzierbar sind, könne von keinem Computer gesagt werden, er habe Geist, und könne das gleiche ebensowenig von einem System behauptet werden, das Hirnfunktionen nachbilde. Allerdings könne es sinnvoll sein, einen „methodischen Funktionalismus“ zu vertreten, nach dem der Geist als Symbolverarbeitungsmaschine anzusehen sei, wenn es darum gehe, eine Theorie des Geistes im Rahmen der psychologischen Forschung aufzustellen (A25, A28).

Wender greift in seinem Beitrag⁸ „**Computersimulation als eine Methode der Psychologie**“ genau das auf, was Kemmerling im Kontext des methodischen Funktionalismus als Simulationsthese erwähnt. Diese These besagt, daß eine Theorie des Geistes durch die Psychologie derart präzise formuliert werden kann, so daß diese Theorie auch als Computerprogramm formulierbar ist (A27, S.25). Wender zeigt, daß die Simulationsthese wie folgt mit einem bestimmten Menschenbild der kognitiven Psychologie einen Engpaß der Theoriebildung darstellt, da sie die Fülle des empirischen Materials mit Hilfe „älterer“ Beschreibungsmittel nicht bewältigen kann (A2). Entweder man ziehe hieraus den Schluß, menschliches Verhalten sei nicht formal beschreibbar, oder aber man versuche es mit anderen Beschreibungsmitteln (A7). Als ein alternatives Beschreibungsmittel für „kognitive Prozesse“ könnten

⁷S.16-S.27, 31 Abs.

⁸S.94-S.104, 33 Abs.

„Rechnerprogramme“, und zwar vor allem KI-Programme (A1), angesehen werden, wenn man voraussetzt, daß „kognitive Vorgänge“ informationsverarbeitende Prozesse sind (A11). Dabei erklärt Wender, daß die theoretische Annahme, daß sich kognitive Prozesse, so wie sie durch den Menschen abgewickelt werden, als Programme simulieren lassen, zum „impliziten Menschenbild der kognitiven Psychologie oder Kognitionswissenschaft“ gehören (A30). Nebenbei merkt er jedoch auch an, daß sich diese Auffassung nur auf Teile der kognitiven Prozesse bezieht. Und alle beispielhaften Programme, die er nennt, die als Simulationen geistiger Phänomene interpretiert werden können, haben nichts mit Empfindung, Wille oder Selbstwahrnehmung zu tun.

Während Wender für die kognitive Psychologie ein Menschenbild skizziert, in dem einige kognitive Prozesse als KI-Programme verstanden werden können, thematisiert **Strube** in seinem Beitrag⁹ „**Die Rolle psychologischer Konzepte in der Künstlichen Intelligenz**“ einen Vorgang in umgekehrter Richtung. Er argumentiert, solange das Ziel der KI die Konstruktion menschlicher Intelligenz ist, müssen in der KI-Forschung psychologische Konzepte verwendet werden, da die Psychologie die Wissenschaft zur Untersuchung der menschlichen Intelligenz ist (A10, A19). So sind etwa unter Berücksichtigung der funktionalistischen Position für die Kognitionswissenschaft bei geistigen Phänomenen „begriffliche Kategorien“ erforderlich, die der „menschlichen und der maschinellen Intelligenz vorgängig sind“ (A11). Für den Zusammenhang Menschenbild und KI zeigt sich hier ein weiterer Aspekt, den Strube nicht explizit ausdrückt. Es könne sein, daß die KI-Forschung gezwungen sei, aufgrund ihres Forschungsanspruches über die psychologischen Kategorien an einem Menschenbild zu partizipieren, in dessen Rahmen geistige Phänomene nicht KI-technisch zu realisieren seien. In dieser Perspektive hätten die Stichworte Geist oder menschliche Intelligenz ausschließlich stimulierende Funktion für den sogenannten KI-Bereich der Informatik (A20). Allerdings weist Strube ausdrücklich darauf hin, daß sich die Kognitionswissenschaft die Aufgabe gestellt hat, „subjektives Erleben und objektiv beobachtbares Verhalten“ zu korrelieren (A13), womit wiederum der Reichweite technischer Simulationen menschlichen Verhaltens Vorschub geleistet wird. Tatsächlich schließt Strube die Möglichkeit, intelligentes Verhalten oder „Handeln“ technisch irgendwann realisieren zu können, nicht aus (A16, A20), hofft aber, daß es den Ingenieuren nicht darum geht, einen Menschen zu konstruieren (A19).

Während Kemmerling den Computer als Menschenbild beziehungsweise als ein Bild des Geistes disqualifiziert, da durch ihn wesentliche geistige Komponenten (etwa Emotionen) nicht erklärt werden können, und Schlachetzki (unzulängliche Tests) und Wender („Holzbein“-Simulationen) die technischen Möglichkeiten anzweifeln, ohne für eine derartige kritische Sicht

⁹S.83-S.93, 20 Abs.

ein Menschenbild bemühen zu müssen, taucht nun bei Strube die Unterscheidung in ein Menschenbild der KI und ein Menschenbild der KI-Entwickler und KI-Ingenieure auf. Die KI-Forschung ist darauf festgelegt, ein einseitiges, nämlich technisch umsetzbares Menschenbild zu entwickeln, jedoch nicht der Ingenieur, in dessen Menschenbild auch technisch irrelevante wie ethische Belange Geltung haben sollten.

In ihrem Beitrag¹⁰ „**Künstliche Intelligenz: Eine Phänomenologische Kritik**“ attestieren **Becker/Lischka** den Konzepten der KI zunächst das Menschenbild, das auch von Capurro/Seetzen als epistemisch einseitiges Menschenbild der KI genannt wird. Becker/Lischka kritisieren in erster Linie aber nicht die Einseitigkeit dieses Menschenbildes, sondern dessen epistemischen Wert. Das KI-Menschenbild Descartes'scher Tradition leugne schlicht die Beziehung zwischen Geist und Körper, die durch die Leiblichkeit vermittelt (A6) und für alle Erkenntnisprozesse konstitutiv sei. Dieser Kritik unterliegt auch die funktionalistische Position, da sie den Leib-Seele- oder Subjekt-Objekt-Dualismus zugunsten einer naturwissenschaftlichen Erkenntnis löst. Nicht nur die „klassische“ KI (A9), auch neuere kritischere Ansätze (A10) scheiterten, indem sie die dialektische Beziehung zwischen Subjekt und Welt übersähen, die durch die Leiblichkeit hergestellt sei. Ein anderes Erkenntnis-konzept, das dem Menschen der phänomenologischen Tradition entnommen sei, könne allerdings nur um den Preis der Omnipotenz naturwissenschaftlicher Erkenntnis vertreten werden. Sehr wichtig ist Becker/Lischka die Einschätzung der sozialen Folgen einer KI, die dem kritisierten Menschenbild entspricht. Danach gehöre KI in die Reihe der sozialen Ausgrenzungssysteme, d.h. strebe Standardisierung, Entindividualisierung und die Verdrängung alles Irrationalen an (A1). Indem nur zugelassen werde, was in das „technische Modell“ passe (A12), trete KI als „materialisierte Ordnungsfunktion“ auf und müsse wie alle Versuche, eine allgemeine Ordnung herzustellen, an dem „besonderen Verhältnis von Welt und Subjekt“ scheitern (A11). Nach der Auffassung von Becker/Lischka ist KI mit ihren gegenwärtigen Strategien zum Scheitern verurteilt, und auch die Gefahr einer grotesken „KI-Ordnung“ besteht im Grunde nicht, da ihr ein unzutreffendes Menschenbild zugrunde gelegt ist. Von hier aus eröffnet sich auch die Chance, KI-Technik im Rahmen eines anderen Menschenbildes sinnvoll als „Sozial-, Human-, und Körpertechnik“ zu entwickeln (A17).

Eine positive Aussicht für KI, die auf ein anderes Menschenbild gestützt ist, eröffnet auch **Müller** in seinem Beitrag¹¹ „**KI und Menschenbild im Unternehmen**“. Müller geht davon aus, daß das Menschenbild moderner Unternehmensführung nicht mit dem Menschenbild der „klassischen KI“ vereinbar ist (A1). Dabei komme es wesentlich darauf an, daß das

¹⁰S.38-S.48, 21 Abs.

¹¹S.105-S.124, 48 Abs.

traditionelle KI-Menschenbild nahelege, es könnten autonome, intelligente Systeme konstruiert werden, die irgendwann abgegrenzte Einheiten in hierarchisch strukturierten Unternehmen substituieren sollen. Auf der Grundlage eines solchen Menschenbildes sei der Einsatz von KI-Produkten in Unternehmen unproduktiv (A6). Demgegenüber vermeide das Menschenbild der modernen Unternehmensführung eine Unterschätzung der sozialen Dimension (A46), indem es das Unternehmen als einen Organismus darstelle, der sein Überleben zu organisieren habe, wofür Flexibilität und nicht standardisiertes Handeln notwendig seien (A11). In der Perspektive dieses Menschenbildes werde das rein technische von einem soziotechnischen Systemverständnis abgelöst (A28), das Konzept einer individuellen oder von „Trägereinheiten“ monopolisierten Intelligenz müsse durch das Konzept einer „verteilten Intelligenz“ ersetzt werden (A23, A25). KI-Systeme seien im Sinne der verteilten Intelligenz nach der Idee des „Computers als Medium“ in die sozialen Strukturen des Unternehmens zu integrieren (A48). Und Müller betont, daß dieses Menschenbild nur mit wirklichen soziotechnischen Umsetzungen verträglich ist und daß es nicht um eine „moralische Schiene der 'menschengerechten Technik'“ geht (A47).

Becker/Lischka und Müller fügen der Position von Capurro/Seetzen und der von Kemmerling, Schlachetzki, Wender, Cremers/Eder/Hinze und Strube sowie der von Eckmiller eine weitere hinzu. Nach ihrer Position ist KI, solange sie nicht unter den Bedingungen eines anderen Menschenbildes entwickelt wird, auf der Basis des kritisierten Menschenbildes als Technik zum Scheitern bestimmt.

Haberbeck teilt in seinem Beitrag¹² „**KI-Perspektiven der Anwendung und Technologiefolgenabschätzung**“ in gewisser Hinsicht die Auffassung Müllers. Er nennt das Bild des „Computers als Medium“ den „Verwendungszusammenhang“ (A3), der für die Entwicklung und den Einsatz von KI im Vordergrund stehen soll. Aus unternehmerischer Sicht plädiert er für die rechtzeitige Einbeziehung der KI in den Verwendungszusammenhang, um technologischen Fehlentwicklungen vorzubeugen (A15). Gegenüber Müller thematisiert Haberbeck aber vor allem die Verantwortung bei der Entwicklung von KI-Produkten. Wobei er die Schwierigkeit sieht, die Verantwortungsfrage isoliert für KI zu stellen (etwa im Zuge einer Technologiefolgenabschätzung für ein KI-Produkt), da einzelne KI-Komponenten im Verwendungszusammenhang aufgehen (A13), und weil letztlich auch die „Leitbilder“ für KI-Entwicklungen lediglich den Leitbildern des Verwendungszusammenhangs entlehnt sind (A22). Diese Position legt nahe, die Diskussion um ein Menschenbild der KI in eine Diskussion um die Menschenbilder der Verwendungszusammenhänge zu überführen. Haberbeck stellt jedoch keinen Zusammenhang zwischen Menschenbildern und dem Problem der Verantwortung her.

¹²S.125-S.135, 24 Abs.

Ohne ein bestimmtes Menschenbild der KI oder Verwendungszusammenhänge zu kritisieren, fordert er, um die „Verantwortbarkeit entsprechender Anwendungen gewährleisten“ zu können, Technologiefolgenabschätzungen, die die Industrie, Forschungseinrichtungen und die Öffentlichkeit mit einbeziehen (A24).

Kremer ist mit seinem Beitrag¹³ „**Informationstechnik im gesellschaftlichen System**“ in der Nachbarschaft von Haberbeck anzusiedeln. Auch bei ihm findet sich eine konzeptionelle Trennung zwischen der Entwicklung von KI, die sich an einem bestimmten Menschenbild orientiert, und der Verantwortbarkeit des Einsatzes beziehungsweise der Entwicklung von KI. In diesem Zusammenhang muß Technik als Komponente gesellschaftlicher Systeme verstanden werden (A6), und das heißt - hier begegnet Kremer der Position von Müller -, daß das Bild einer KI als reiner Technik durch das Bild soziotechnischer Systeme zu ersetzen ist (A12). Um Resultate eines soziotechnischen Systems, die „unseren Interessen“ zuwiderlaufen, zu vermeiden, sei Technologiefolgenabschätzung erforderlich (A14) beziehungsweise seien gesellschaftliche Regelmechanismen auszubauen, die die Verantwortung für technische Produkte wahrnehmen könnten. Genauso wie von Seetzen/Capurro, Schlachetzki und Haberbeck werden von Kremer etablierte Organisationen, Institutionen, aber auch Initiativen als Verantwortungsträger genannt (A16). Zusätzlich räumt Kremer den Ingenieuren die besondere Funktion ein, vom rein technischen Standpunkt aus Bewertungsgrundlagen für verantwortliche Entscheidungen aufzuzeigen (A19, A20).

Schreiber plädiert mit seinem Beitrag¹⁴ „**KI-Produkte und Verantwortung**“ ebenfalls dafür, den KI-Techniken und -Produkten über die Frage der Verantwortung beizukommen. Er problematisiert nicht das Menschenbild der KI, sondern die besonderen Schwierigkeiten, denen der verantwortungsvolle Umgang mit KI ausgesetzt ist. Die entwickelten gesellschaftlichen Mechanismen seien durch die Komplexität der modernen technischen Systeme teilweise überholt. Zum Beispiel das Prinzip der Haftbarkeit sei an eine juristische Person geknüpft, und es fiele schwer, einer Maschine über den Status einer juristischen Person Verantwortung und Haftbarkeit zu übertragen (A4, A5). Unabhängig von rein kriminellen Mißbräuchen bergen etwa fehlerhafte Expertensysteme ein anderes Gefahrenpotential in sich als ein fehlerhaftes Fachbuch (A9). Im Sinne von Seetzen/Capurro hält auch Schreiber „kollektive Verantwortung (Haftung)“ für geboten (A11). Darüberhinaus gibt Schreiber in seinem Beitrag viel Raum für den Vorschlag, „Verantwortungsfunktionen“ in KI-Produkte einzubauen (A12, A13, A16-A25). Damit setzt Schreiber einen neuen Akzent, da er das Menschenbild der KI

¹³S.143-S.155, 24 Abs.

¹⁴S.171-S.179, 25 Abs.

nicht nur lediglich nicht kritisiert, sondern sogar bereit ist, die Verantwortungsfrage teilweise an die KI weiterzugeben, d.h. in ein KI-Menschenbild zu integrieren.

Die Auffassung, die aus dem Beitrag¹⁵ „**Potentielle Gefahren von KI-Systemen**“ von **Röpke** hervorgeht, beansprucht auch keine Kritik des Menschenbildes, das die KI-Forschung leitet, sondern zielt wiederum auf den besonders verantwortungsbewußten Umgang mit KI ab. Röpke zählt einige potentielle Gefahren auf (A5-A10) und bemerkt, daß kein KI-System mit einem Null-Risiko funktionieren kann (A12).

Im Gegensatz zu der Perspektive, die Röpke, Schreiber, Kremeier und Haberbeck einnehmen, in der ein Menschenbild der KI keine wesentliche Rolle spielt, sondern alles entscheidend von den Verantwortungskonzepten der Gesellschaft abhängt, stellt **Krämer** in ihrem Beitrag¹⁶ „**Eine weitere kopernikanische Wende?**“ einen Zusammenhang zwischen dem Menschenbild der KI und der Verantwortungsfrage dar. Sie schreibt: „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.“ (A28, S.36) Obwohl Descartes als Mitbegründer der rationalistischen Tradition oft als Vater der KI verstanden werde, sei es gerade der Geistbegriff des Rationalismus, der durch das KI-Konzept desavouiert werde. Geist sei für den Rationalismus eine „anthropozentrische Kategorie“ gewesen, die die Subjektivität (A22), letztlich das Konzept von der Person einschließe. Mit dem Modell der KI vom Geist als einer Informationsverarbeitungsmaschine werde der rationalistische Geistbegriff von allen subjektivistischen Unklarheiten oder Unwägbarkeiten bereinigt (A24). D.h. also, indem der Intellekt, der im Rationalismus ein Mittel für moralisches Handeln gewesen sei (A26), von den subjektivistischen Bedingungen gelöst werde, entstehe im Windschatten eines KI-orientierten Menschenbildes ein Geistbegriff, der das Moralische nicht mehr notwendig mit einschließe (A27). Das würde für sämtliche Positionen, die eine Betonung der Verantwortung im Umgang mit KI fordern, bedeuten, besonders das Menschenbild zu hinterfragen, das die KI reproduziert, um an dessen Stelle ein Menschenbild zu setzen, das von einer wesentlichen Beziehung zwischen Geist und Verantwortung ausgeht, und das der KI-Forschung eine andere Orientierung gäbe.

Unter den wesentlichen Fragen, die **Görz et al.** in ihrem Beitrag¹⁷ „**Mögliche Auswirkungen einer entwickelten KI auf Arbeits- und Lebenswelt**“ in einem Zukunftsszenario mit vollentwickelter KI-Technik berühren, taucht auch das Problem der Verantwortbarkeit auf. Sie

¹⁵S.180-S.185, 15 Abs.

¹⁶S.28-S.37, 28 Abs.

¹⁷S.156-S.170, 35 Abs.

bringen dabei einen weiteren Aspekt zur Sprache, nämlich die Abhängigkeit des Menschen von Maschinen, in die er sich begeben könnte, wenn er den Maschinen Verantwortung überträgt (A14). Wie auch Becker/Lischka heben sie die Bedeutung der Leiblichkeit hervor, die den derzeitigen Maschinen als Bezugssystem fehlt, und das konstitutiv für Verantwortung ist (A15). Je konsequenter die Darstellung des Szenarios ausfällt, desto deutlicher zeigt sich, daß mit der Ausdifferenzierung der KI-Technik auch die Schwierigkeiten bei der Technikbewertung zunehmen (A20-A24). Zwar steht in diesem Beitrag nicht das Menschenbild der KI zur Disposition, dafür aber das Menschenbild, das wir von uns als Menschen als Typ unter den KI-Systemen haben. Denn offensichtlich setzt unsere Ethik für das Verhältnis Mensch und Umwelt eine „Sonderstellung“ des Menschen voraus: „Es ist vorstellbar, daß ethische Dimensionen ihren ausschließlichen Bezug auf menschliches Verhalten in dem Maße verlieren, wie Maschinen humane Fähigkeiten simulieren können.“ (A34, S.169)

4. Folgerungen

Sieht man die Beiträge des Reports durch, ohne sie entlang bestimmter Interessen beurteilen oder systematisieren zu wollen, so eröffnen unterschiedliche Kombinationen der Beiträge sehr verschiedene oder zum Teil auch kühne Perspektiven.

Nimmt man beispielsweise die Position von Schlachetzki auf, nach der allen KI-Techniken gemeinsam ist, daß sie ein mechanistisches Weltbild beziehungsweise Menschenbild realisieren, und liest man seine Einschätzung „Ob trotz aller Erfolge [der mechanistischen KI, GG] das mechanistische Bild vom Menschen, das all dem zugrundeliegt, tatsächlich auch zutrifft, ist eine Frage der Überzeugung...“ (Abs.22, S.80f.), und fügt man dem die Auffassung Schreibers hinzu, „daß entsprechend den Ergebnissen der Analysen möglicher Gefahren vom Hersteller bei solchen [gefährlichen, GG] KI-Produkten *prophylaktisch wirkende Verantwortungsfunktionen* (...) vollständig angelegt sein müssen...“ (Abs. 12, S.175), dann läßt sich hier zusammen mit Eckmillers Konzept vom 'Universal-Plan' (siehe Zitat S.9) eine Perspektive erkennen, in der der Mensch als ein Automat, als Bestandteil des mechanistischen Weltbildes aufgefaßt werden muß, dessen intelligente Fähigkeiten und dessen Verantwortungsfähigkeit konsequenterweise auch als KI-Funktionen technisch umsetzbar sein sollten.

Faßt man demgegenüber die Beiträge von Becker/Lischka und Müller zusammen, so eröffnet sich eine gänzlich andere Perspektive. Becker/Lischka kritisieren in einer philosophischen Auseinandersetzung mit KI, daß es sich um Ausgrenzungspraktiken handelt: „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." (Abs.12, S.44) Da aber „die menschliche Praxis durch ein nicht zu vernachlässigendes Maß an Ungeregeltem gekennzeichnet ist" (Abs.14, S. 44), müßte an technischen Entwicklungen gearbeitet werden, zu deren Voraussetzungen ein nicht mechanistisches Menschenbild gehört. Hier mündet die philosophische Auseinandersetzung in die Einschätzung, daß nur durch eine Einschränkung des mechanistischen Menschenbildes technische Konzepte entwickelt werden können, die das Scheitern gegenwärtiger KI-Technik überwinden könnten. Dieser Argumentationsrichtung folgt Müller aus unternehmerischer Sicht: „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." (Abs.47, S.122) Das geforderte Systemverständnis, das zum Erfolg eines industriellen Produktionssystems beitragen soll, widerspricht einem mechanistischen Weltverständnis.

Es können durch entsprechende Kombinationen der Beiträge auch sehr kontroverse Perspektiven aufgezeigt werden. Etwa die Beiträge Becker/Lischka und Capurro/Seetzen und Schreiber dürften zu einem Streit um Ausgrenzungstechnik und Verantwortung führen. Obgleich Becker/Lischka nicht auf Verantwortung abzielen, worauf sich Capurro/Seetzen konzentrieren, darf man vermuten, daß diese beiden Beiträge sich nicht gegenseitig die Voraussetzungen streitig machen. Dagegen sieht es so aus, daß Schreiber KI-Produkte und -Produktionen nicht als Ausgrenzungstechnik betrachten kann, und es daher auch für möglich hält, Verantwortung auf KI-Produkte zu übertragen. Entweder Capurro/Seetzen und Schreiber haben völlig unterschiedliche Vorstellungen darüber, was Verantwortung sein könnte, oder sie meinen in ihren Beiträgen, die ja zu **einem** Diskursprojekt gehören, mit dem gleichen Ausdruck „Verantwortung" völlig verschiedene Sachverhalte. In letzterem Fall, würde man erwarten, dürfte ein fortgesetzter „Diskurs" hierüber schnelle Einigung erzielen.

Für jede Kombination der Beiträge des Reports, die eine kontroverse Perspektive eröffnet, scheint die Diskrepanz in Begriffsverwendungen beziehungsweise im Begriffsverständnis symptomatisch zu sein. Die Gründe für eine derartige Diskrepanz sind nicht offensichtlich. Vielleicht bringt die Systematik, die die Beiträge unter bestimmten Gesichtspunkten kombiniert, auch einiges über die Ursachen für die sehr divergierenden Begriffsverwendungen und -auffassungen ans Licht.

Bevor wir uns die Systematik zunutze machen, um die Konflikte zwischen den Beiträgen zu erhellen, möchte ich darauf hinweisen, daß die Systematik, die einer Durchsicht des Reports 17 zu verdanken ist, einer inhaltlichen Strukturierung ähnlich ist, die auch in der Diskussion zur KI außerhalb des Diskursprojekts zu finden ist. So gibt es (I) eine KI-technische Diskussion. Auch diese Diskussion erlaubt Erweiterungen in einer weniger technischen Richtung, etwa auf der Grundlage einer Auseinandersetzung zwischen Konnektionismus und Symbolverarbeitung. Aber natürlich gibt es reichlich Publikationen, in denen es um KI-Technik geht. Ein bekannteres Beispiel für diesen (erweiterten) Diskussionstyp ist das Buch „Understanding Computers and Cognition“ (1986) von T. Winograd und F. Flores. Dann gibt es (II) die „erkenntnistheoretische“ Diskussion, die zum Beispiel in der Kognitionswissenschaft geführt wird. Für diese Diskussion sind technische Gesichtspunkte keinesfalls hinreichend. In extremen Fällen beteiligen sich Autoren an dieser Diskussion, indem sie zunächst alle denkbaren technischen Erfolge unterstellen, um dann „sehr philosophisch“ über erkenntnistheoretische Implikationen der KI zu argumentieren. Für diesen Diskussionstyp ist das Buch „What Computers Can't Do“ (1972) von H. Dreyfus beispielhaft.

Eine weitere Diskussion (III) ist die ethische, die sich mit dem Verantwortungsaspekt der vorliegenden Systematik verbinden läßt. Es hat den Anschein, als würden Teilnehmer der ethischen Diskussion immer wieder versuchen, Beiträge aus den Diskussionen (I) und (II) zu instrumentalisieren. Dabei sind aber in (III) weder technische noch (erkenntnis-)theoretische Argumente hinreichend, obwohl sie dort unter Umständen große Überzeugungskraft haben könnten. Zur Illustration dieser Diskussion kann etwa ein „Streitgespräch“ zwischen J. Weizenbaum und K. Haefner herangezogen werden (Der Spiegel 10/1987, S.92-112).

Die Frage ist nun einerseits, in welcher Hinsicht die Strukturierung in (I) - (III) für die KI-Diskussion bezeichnend ist, und andererseits, welche Schlüsse daraus zu ziehen sind, daß die Autoren des Reports 17 in der Systematik (a) - c) je eine bestimmte Stelle einnehmen. Zunächst erscheint es sehr einleuchtend, daß es davon abhängt, in welcher Art der Diskussion ich mich befinde, wie ein bestimmtes Argument bewertet wird. So ist es, wenn ich in eine Diskussion vom Typ (II) verwickelt bin und gegen die technische Simulation von emotionalen Zuständen argumentiere, wenig relevant, ein technisches System zu berücksichtigen, das sich in irgendeiner Hinsicht so verhält, als simuliere es Emotionen. Ein technisches Faktum muß in einer solchen Diskussion überhaupt kein Gewicht haben. Dies verhält sich erst dann anders, wenn in einer Diskussion vom Typ (II) Ergebnisse gewonnen werden, die die Möglichkeit einer technischen Simulation von emotionalen Zuständen nicht ausschließen. Wahrscheinlich ist es sehr selten der Fall, eine Diskussion vom Typ (I), (II) oder (III) vorzufinden und fast immer handelt es sich um „Mischdiskussionen“. Falls jedoch

die Einschätzung der Argumente vom jeweiligen Typ der Diskussion abhängt, liegt die Vermutung nahe, daß eine „Mischdiskussion“ eine wesentliche Ursache für Kontroversen ist, in denen wichtige Begriffe völlig unterschiedlich gebraucht beziehungsweise verstanden werden.

Versuchen wir, unsere Systematik auf den Report 17 anzuwenden, um die Begriffsdiskrepanzen, die für die kontroversen Beiträge symptomatisch sind, transparenter zu machen. Zum Beispiel die Beiträge von Krämer und Strube sind in der Systematik an verschiedenen Stellen angesiedelt (siehe die Abbildung, S.5). Bei Krämer spielt der Verantwortungsbegriff eine besondere Rolle, während Strube nur marginal auf Verantwortung Bezug nimmt, dabei aber offensichtlich einen anderen Akzent setzt.

Krämer arbeitet in ihrem Beitrag heraus, daß in der KI einerseits ein Aspekt des Rationalismus in der cartesischen Tradition (der der Geistestechnik) ausgebildet wurde, und andererseits ein Aspekt des gleichen rationalistischen Projekts (der des Willens, der Subjektivität) ausgeklammert wurde. So weist sie darauf hin, daß die erkenntnistheoretischen Leistungen des cartesischen Rationalismus an ein „moralisch verantwortliches Subjekt“ gebunden sind. Sie schreibt: „Die Bezugnahme auf eine der Verantwortung fähigen und zur Verantwortung pflichtigen Person wird zum konstitutiven Element des cartesischen Begriffes vom Geist.“ (Abs.20, S.33) Hier ist der Akzent der, daß die Geistestechniken über das Konzept des willentlichen Subjektes an einen Begriff der Verantwortung gebunden sind, und zwar derart, daß von hier aus Geist gar nicht ohne einen Begriff der Verantwortung erklärt werden kann.

Strube hingegen, der in seinem Beitrag davon ausgeht, daß eine Geistestechnik sinnvollerweise psychologische Konzepte berücksichtigen sollte, da die Psychologie die Erkenntnisse über die kognitiven Fähigkeiten des Menschen liefere, behauptet, daß erst die Konstruktion sehr vielseitiger beziehungsweise flexibel verwendbarer KI-Systeme die Bedeutung der Psychologisierung der KI zeige. „Und so werden es letztlich anspruchsvolle Einsatzkonzepte für intelligente Maschinen sein, die eine Hinwendung zur kognitiv adäquaten und damit immer auch aufwendigen Lösungen erzwingen werden. Dies gilt nicht erst für ein (hoffentlich nicht nur abstrakt gemeintes) Ziel wie das, einen künstlichen Menschen bauen zu wollen.“ (Abs.19, S.91)

Gehen wir davon aus, daß Strube seinen in Klammern gesetzten Einschub seinem Verantwortungsverständnis verdankt, dann müßte sich der Verantwortungsbegriff unterscheiden, den Krämer darstellt. Die Pointe bei Krämer ist gerade die, daß es der cartesische Begriff vom Geist nicht gestattet, ein künstliches System anzunehmen, das über die kognitiven Fähigkeiten

des Menschen verfügt ohne verantwortungsfähig sein zu können. Für Strube, der Verantwortung nicht zu den Voraussetzungen eines KI-Systems zählt, ist Verantwortung hingegen etwas, das dem KI-Entwickler zuzurechnen ist. Hier ist Verantwortung nicht konstitutiv für kognitives Verhalten, sondern etwa an die funktionelle Rolle eines kognitiven Systems in der Gesellschaft gebunden.

(Weniger künstlich würde der Kontrast zweier Bezüge auf Verantwortung wirken, wenn man zum Beispiel den Verantwortungsbegriff von Schreiber (siehe oben) dem von Krämer gegenüberstellt.)

Es läßt sich allein aufgrund der Beiträge auch nicht klären, ob es hier zum Beispiel zwei unterschiedliche Auffassungen zur Verantwortung sind, die Krämer und Strube trennen, oder ob es sich lediglich um ein oberflächliches Mißverständnis handelt. Beziehen wir uns nun auf die Einordnung in die Systematik, so gilt für den Beitrag von Strube, der in die Gruppe der Beiträge gehört, die KI als Technik thematisieren, daß es in ihm kein Argument mit Gewicht geben kann, das unabhängig von den KI-technischen Entwicklungen formuliert wird und „anti-technische“ Konsequenzen (gegenwärtige Systementwicklungen) keinerlei Einfluß auf die Argumentationsstrategien haben.

Es sieht also so aus, als sollten sinnvollerweise nur die Beiträge verglichen werden, die in der Systematik an der gleichen Position sind und als hätte der Vergleich unter jenem Gesichtspunkt zu erfolgen, durch den die Beiträge der gleichen Position zugeordnet worden sind. Jeder andere Vergleich mag spektakulärer, aber nicht ergiebiger sein.

Sind die Leitvorstellungen der KI verantwortbar?

Dr.-Ing. J. Seetzen und G. Grube

"Die Diskussion um die Möglichkeiten der KI ist in den letzten zehn bis fünfzehn Jahren in Amerika sehr viel heftiger geführt worden als in Europa. Dabei ging es im wesentlichen um die Frage, was man von KI-Lösungen erwarten kann. Diese Diskussion schließt an die ältere an, was man von Computern erwarten kann, und diese wiederum ist eine Fortsetzung der Diskussion um die Folgen der Kybernetik, die in den fünfziger und sechziger Jahren geführt wurde.

Es wäre nützlich gewesen, in unserer Arbeit die amerikanische Diskussion stärker in Betracht zu ziehen. Wir wollen dies ein kleines Stück weit nachholen. Es zeigt sich aber, daß es eine Tiefenschicht gibt, die auch in Beiträgen unseres Berichtes durchscheint, nämlich die alte Frage, inwieweit unser Verhalten und damit auch die Verantwortung auf Wissen gegründet werden kann. Mit anderen Worten, im Blick auf die Verantwortbarkeit von Leitvorstellungen, wie sie von KI-Protagonisten geäußert werden, muß viel mehr auf die über das "Wissen" hinausgehende Funktion von Informationen hingewiesen werden, als das bisher geschehen ist (Seetzen 1992).

Wir zitieren zunächst zwei Äußerungen von amerikanischen KI-Spezialisten. Marvin Minsky (1992) hat gesagt "Ich finde, Affen sind Computer, und Menschen sind bessere Computer. Meiner Meinung nach sollten die Leute eher beleidigt sein, eine Ansammlung von Chemikalien genannt zu werden. Es ist gut, ein Computer zu sein. Wenn wir verstehen, wie das Gehirn als Computer funktioniert, dann können wir ewig leben, indem wir Gehirnprozesse kopieren. Der Tod des Gehirns ist nur ein chemischer Unfall. Durch echte KI können wir den Tod besiegen. Die Leute sollten ihr Geld statt der Kirche lieber der Forschung geben, dann könnten wir ihre Seele wirklich retten."

Ein zweites etwas längeres Zitat stammt von Hans Moravec aus dem Prolog (1988) zu seinem Buch "Mind Children": "Heute sind unsere Maschinen noch einfache Schöpfungen, die der elterlichen Sorge und Pflege eines Neugeborenen bedürfen und kaum das Attribut 'intelligent' verdienen. Aber im nächsten Jahrhundert werden sie zu Wesen heranreifen, die so komplex

sind wie wir und schließlich zu etwas werden, das alles, was wir kennen, überschreitet - aber auf das wir stolz sein können, wenn sie sich untereinander als unsere Abkömmlinge bezeichnen. Losgelöst vom schwerfälligen Fortgang der biologischen Evolution, werden die Kinder unseres Geistes frei sein, um außerordentlichen sowie fundamentalen Herausforderungen im größeren Universum gewachsen zu sein. Wir Menschen werden für einige Zeit Nutzen von ihrer Arbeit haben, aber früher oder später werden sie wie natürliche Kinder ihr eigenes Glück suchen, während ihre alt gewordenen Eltern still dahin gehen. Es braucht bei dieser Übernahme der Fackel nur wenig verloren zu gehen - es wird in der Entscheidungsmacht unserer künstlichen Nachkommen liegen, und ihr Nutzen sein, ziemlich alles über uns in der Erinnerung zu behalten, vielleicht sogar das Arbeiten des Geistes von menschlichen Individuen im einzelnen."

Man könnte diese zynischen oder pathetischen Äußerungen ablehnend zur Kenntnis nehmen als Ausdruck von Geistern, die zu viel "Science Fiction" konsumiert haben. Aber in abgemilderter Form geistern derartige Vorstellungen in dem von amerikanischen KI-Verfechtern inspirierten, vordergründig gescheiterten japanischen "Fifth Generation Computer Program" und dessen Nachfolgeprogramm, dem "Real World Computing Program" völlig unangefochten herum und Moravec's "genetic takeover" wird unkritisiert in unserem Bericht von sieben Autoren gemeinsam zitiert. Die Faszination der KI-Phantasien als Leitvorstellungen ist also ernster zu nehmen.

Die früheste Kritik an den Entwicklungen der KI in Amerika stammt von Joseph Weizenbaum (1978). Man kann aus der Rückschau sagen, daß dies die Stimme eines betroffenen Humanisten ist. Die Kritik Weizenbaums ist aber stärker in Europa als in Amerika aufgenommen worden.

Dagegen hat sich eine spezielle Kritik gegen tatsächlich oder vermeintlich übertriebene Vorstellungen der Möglichkeiten der KI in Amerika herausgebildet, die besonders von John Searle (1986), den Brüdern Hubert und Stuart Dreyfus (1987), Terry Winograd und Fernando Flores (1987) vorgetragen worden ist. Diese Kritik hat im wesentlichen zwei Ansätze verfolgt. Die eine Form der Kritik ist praktischer Art. Das heißt, es wird kritisiert, daß die Ansprüche von KI-Entwicklern als Realität dargestellt werden, diese vermeintliche Realität aber nicht auf ihre Konsequenzen hin betrachtet wird. Zum Beispiel, was würde es wirtschaftlich bedeuten, wenn durch Robotik (einem Teilgebiet der KI) Produktionen hochgradig automatisiert werden, also menschliche Arbeit in erheblichem Maße überflüssig wird. Letztlich führt diese praktische Kritik dazu, menschliche Maßstäbe oder Leitvorstellungen in ihrem eigenen

Recht zu beachten, auch wenn es durchaus Perspektiven der Informatik oder KI gibt, die diese außer Kraft setzen könnten.

Eine weitere Form der praktischen Kritik stellt den Systemzusammenhang zwischen KI-Produkt und menschlicher Handlung in den Vordergrund und fragt nach der Verantwortbarkeit solcher anthropotechnischen Prozesse. Denn letztlich ist es ethisch sinnlos, dem KI-Produkt allein Verantwortung (auch im rechtlichen Sinn) zuzuweisen.

Die andere Richtung der Kritik ist theoretischer Natur. Sie bemüht sich von vornherein, Unmöglichkeiten der Ansprüche von KI-Entwicklungen festzustellen. Hierzu zählen:

- 1) Alles Wissen ist explizierbar und kann in einem Formalismus ausgedrückt werden. (Erkenntnistheoretische These).
- 2) Alles Seiende besteht aus einer Menge von Tatsachen, die allesamt logisch voneinander unabhängig sind. (Ontologische These).
- 3) Denken ist Informationsverarbeitung auf der Grundlage regelgeleiteter Symbol-manipulationen. (Symbolistische These).
- 4) Das entscheidende Kriterium für Kognition ist erfolgreiche Abbildung einer Außenwelt. (Repräsentationsthese).
- 5) Auf der unteren Ebene der Kognition sind es nicht Symbole, deren Bedeutungen zu Gesamtbedeutungen verknüpft werden, sondern bedeutungslose Einheiten, die auf der oberen Ebene ein Netzwerk bilden, das sich intelligent verhält. (Emergenzthese).

Die Kritiker versuchen nachzuweisen, daß diese Thesen der KI nicht haltbar sind. Von ihrem Standpunkt aus haben sie recht. Eine Einigung in diesem 'Streit der Fakultäten' wäre nur zu erreichen, wenn jede Seite sich auf die Semantik der anderen einlassen würde. Aber dies ist in der amerikanischen wie in der europäischen Wissenschaftslandschaft nicht üblich, und so erscheinen die semantischen Barrieren als nur sehr schwer übersteigbar, was auch die Erfahrung unseres Ausschusses ist.

Die geistesgeschichtlichen Ursprünge des Denkens, das zu den Thesen der KI geführt hat, gehen weiter zurück.

Es gibt eine Arbeit von 1968 des inzwischen etwas in Vergessenheit geratenen Vorläufers der KI in Deutschland, Karl Steinbuch (1968), "Überlegungen zu einem hypothetischen cognitiven System". Darin hat er ebenfalls ziemlich "science fiction"-haft daran gedacht, "ein technisches System zu entwerfen, das in einer überwiegend unbekanntem Außenwelt seine Funktion erhält und an seinen Auftraggeber Beobachtungen aus dieser Außenwelt ökonomisch (im Sinne des Codierungsauftrages) mitteilt." Die Ausführung dieser Funktion ist weitgehend identisch mit Vorstellungen der KI. Steinbuch war vor allem der Erfinder der "Lernmatrix". Also konnte sein vorgestelltes System aus den Anforderungen der Umwelt lernen. Dies ist nichts anderes als das Paradigma der Neuroinformatik von heute. Das Wichtige der Denkfigur von Steinbuch ist allerdings, daß er klugerweise noch die Nabelschnur des Auftrages von Menschen einbezogen hat. Er dachte also, auch wenn er dies nicht ausgedrückt hat, in anthropotechnischen Systemen, die der menschlichen Verantwortung unterliegen.

In seinem Hauptwerk "Automat und Mensch - über menschliche und maschinelle Intelligenz" hat Karl Steinbuch (1961) bereits vieles von dem vorausgedacht, was wir heute in unserer Diskussion um das "Menschenbild in der Künstlichen Intelligenz" verhandeln. Er schließt dabei an die Gedanken des Begründers der Kybernetik, Norbert Wiener (1963), an. Beide haben sich der Idee des Physikalismus verschrieben. Die Welt sei insgesamt - auch lebende Strukturen - aus den Gesetzen der Physik zu erklären. Beide sind, wenn auch unausgesprochen, Monisten.

Wenn man sich auf den Monismus beruft, wie das auch in unserem Bericht an einer Stelle implizit geschieht, muß man sich der Wurzel dieser Philosophie bewußt sein. Sie geht in ihrer ausgeprägtesten Form auf Ernst Haeckel, den großen Biologen und Verkünder des Gedankens der Evolution, zurück. Haeckel hat in seinem besonders populären Werk "Die Welträtsel" (1908) ein "Glaubensbekenntnis eines Naturforschers als Band zwischen Religion und Wissenschaft" abgelegt. Er beruft sich auf das "Substanzgesetz", einem compositum mixtum von Energieerhaltung und Erhaltung der Materie, das ihm die monistische Basis liefert. Er schreibt: "Das abstrakte große Gesetz der mechanischen Kausalität, für das unser kosmologisches Grundgesetz, das Substanzgesetz, nur ein anderer konkreter Ausdruck ist, beherrscht jetzt das Universum ebenso wie den Menschengest; es ist der sichere, unverrückbare Leitstern geworden, dessen klares Licht uns durch das dunkle Labyrinth der unzähligen einzelnen Erscheinungen den Pfad weist."

Dies ist nach heutiger Einsicht in die Physik schlichter Unsinn. Denn Energieerhaltung und Materieerhaltung sind nicht auf einen Nenner zu bringen. Die Energieerhaltung geht der Ma-

terierhaltung voraus. Wir können dies heute sogar in Experimenten der Hochenergiephysik zeigen. Man sollte also bedachter mit der Idee des Monismus umgehen.

Haeckel war aber ebenso wie Steinbuch vorsichtig, was die ethische Dimension des menschlichen Handelns angeht. Beide haben sich auf die ethische Figur des Kant'schen "kategorischen Imperatives" eingelassen. Sie haben nur beide nicht wahrgenommen, daß sie damit kategorial den Monismus oder den Physikalismus verlassen hatten.

Haeckels Zeitgenosse Friedrich Nietzsche war in dieser Hinsicht konsequenter. Im Aphorismus 335 der "Fröhlichen Wissenschaft", "Hoch die Physik", sagt er am Ende "Wir wollen aber die werden, die wir sind, - die Neuen, die Einmaligen, die Unvergleichbaren, die Sich-selber-Gesetzgebenden, die Sich-selber-Schaffenden! Und dazu müssen wir die besten Lerner und Entdecker alles Gesetzlichen und Notwendigen in der Welt werden: wir müssen Physiker sein, um in jenem Sinne Schöpfer sein zu können, - während alle Wertschätzung und Ideale auf Unkenntnis der Physik oder im Widerspruch mit ihr aufgebaut waren. Und darum: Hoch die Physik! Und höher noch das, was uns zu ihr zwingt, - unsere Redlichkeit!"

Nur war Nietzsche ein schlechter Physiker, der aus der Entdeckung des Satzes von der Erhaltung der Energie durch Julius Robert Mayer, dessen Buch "Die Mechanik der Wärme" er kannte, seine abstruse Leitidee von der ewigen Wiederkehr des Gleichen gewonnen hat. Im Nachlaß Nietzsches findet sich der Satz "Das Gesetz von der Erhaltung der Energie verlangt die ewige Wiederkunft". Es ist von erheblicher Bedeutung, daß Nietzsche dieser Gedanke, der für ihn die Unsterblichkeit einschloß, zu einer Zeit tiefster Selbstzweifel aufgeleuchtet ist.

Es ist kein Wunder, daß die Immoralität des Nihilismus, die in der russischen Literatur ihren Ahnherrn hat, im Verbund mit dem Monismus und dessen Nachfolger im Positivismus und Physikalismus, unser Jahrhundert in eine tiefe Verwirrung und letztlich in ein Elend, das bis heute fortwirkt, gestürzt hat.

Kommen wir zum Abschluß auf unsere Frage zurück, ob die Leitvorstellungen der KI verantwortbar sind. Wir sind überzeugt, daß die eingangs zitierten Leitvorstellungen verantwortungslos sind, und zwar ganz einfach, weil sie im Wortsinne inhuman sind. Sie stehen in einer verheerenden geistesgeschichtlichen Tradition. Daß sie wie bei Nietzsche auf Unsterblichkeit hinauslaufen, läßt psychologisch vermuten, daß sie von tief verängstigten Personen stammen.

Es gibt natürlich nicht nur derartige Leitvorstellungen in der KI. Aber wie wir heute feststellen können, laufen die verantwortbaren Leitvorstellungen in der Regel auf mehr oder weniger

nützliche Hervorbringungen der Informatik hinaus, nach dem bekannten Satz "If it works, it is not AI!". Man könnte nach unserer Auffassung gänzlich auf die Bezeichnung KI verzichten, ohne daß sich damit ein wissenschaftlicher Verlust feststellen ließe. Man gewönne damit einen Teil verlorengegangener wissenschaftlicher Redlichkeit zurück.

Was aber ist zu verantworten? Diese ethische Frage ist leichter gestellt als beantwortet. Keinesfalls können wir das, was für Menschen gut ist, aus dem, was wir positiv wissen, ableiten. Georg Picht (1966) hat in seinem Vortrag "Struktur und Verantwortung der Wissenschaft im 20. Jahrhundert" lange vor Hans Jonas das Problem der kollektiven Verantwortung im Spannungsfeld von Wissenschaft, Wirtschaft und Politik bedacht. Philosophie war im griechischen Sinne immer das Bedenken des Wahren und Guten. Dadurch, daß wir in der europäischen Geistesgeschichte der letzten beiden Jahrhunderte den vollen Akzent auf die Wahrheitssuche gelegt haben, sind wir zu Wissensriesen und zu Weisheitszwergen geworden."

Literatur

- Dreyfus, H.L., Die Grenzen Künstlicher Intelligenz - Was Computer nicht können, Athenäum, Königstein/TS, 1985 (engl. Original 1979)
- Haeckel, E., Die Welträtsel, Gemeinverständliche Studien über monistische Philosophie, Kröners Taschen-
ausgabe, 1908
- Hayman, R., Friedrich Nietzsche, W. Heyne Verlag 1985, S. 309
- Minsky, M., Interview Weltwoche, 1992
- Moravec, H., Mind Children - The Future of Robot and Human Intelligence, Harvard University Press,
Cambridge Mass., London England 1988, S. 1; (Übersetzung: J. Seetzen)
- Newell, A., Simon, H., Computer Science as Empirical Inquiry: Symbols and Search, Communications of the
Association for Computing Machinery, 19; 1976
- Nietzsche, F., Fröhliche Wissenschaft, A. Kröner Verlag, Stuttgart, 1965, S.224
- Picht, G., Struktur und Verantwortung der Wissenschaft im 20. Jahrhundert; in: Der Gott der Philosophen und
die Wissenschaft der Neuzeit, Ernst Klett Verlag Stuttgart, 1966, S.106
- Searle, J.R., Geist, Hirn und Wissenschaft, Surkamp Verlag, Frankfurt a.M., 1986 (engl. Original 1984)
- Seetzen, J., Information, Kommunikation, Organisation - Anmerkungen zur Theorie der Informatik ; in: W. Coy
et al. (Hrsg), Sichtweisen der Informatik, Vieweg 1992, S. 83 - 96
- Steinbuch, K., Automat und Mensch - über menschliche und maschinelle Intelligenz, Springer-Verlag, Berlin -
Göttingen - Heidelberg, 1961
- Steinbuch, K., Überlegungen zu einem hypothetischen cognitiven System, in: S. Moser (Hrsg), Information und
Kommunikation, R. Oldenbourg, München - Wien, 1968, S.53 - 61

- Varela, F. Kognitionswissenschaft - Kognitionstechnik, Suhrkamp Verlag, Frankfurt a.M., 1990, (engl. Original 1988)
- Weizenbaum, J., Die Macht der Computer und die Ohnmacht der Vernunft, Suhrkamp 1978 (engl. Original 1976)
- Wiener, N., Kybernetik, Econ, Düsseldorf - Wien, 1963, S.81 und S.192
- Winograd, T., Flores, F., Understanding Computers and Cognition, Addison Wesley Publ. Co., 1987, (deutsch erschienen: Rotbuch Verlag, Berlin 1989)

Müssen die bisherigen Positionen der KI überdacht werden?*

Prof. Dr. J. Siekmann

Ausgangspunkt für den Vortrag von Siekmann ist ein Gutachten für den BMFT von ihm aus dem Jahre 1984 über die allgemeine Lage der KI, in dem er kurzfristige, mittel- und langfristige Prognosen über die künftige Entwicklung der KI gestellt hatte. Das Gutachten zeichnete sich durch eine differenzierte Betrachtung der einzelnen Teilgebiete der KI aus.

Das Bestreben des Vortrags ist es, anhand der damaligen Prognosen festzustellen, wie die einzelnen Teilgebiete der KI sich tatsächlich entwickelt haben und die Gründe für falsche Vorhersagen offenzulegen. Bei solchen Prognosen zeigt sich ein interessanter Trend: Oft werden kurzfristige Entwicklungen in ihrer Bedeutung für gesellschaftliche und wirtschaftliche Zusammenhänge überschätzt, während langfristige Trends offensichtlich in ihrer Tragweite oftmals verkannt werden. Letzteres sollte allerdings nicht dazu führen, daß Wissenschaftler aus Furcht vor falschen Vorhersagen auf Prognosen verzichten und somit die Prognostik und auch die Entscheidungsfindung ausschließlich in den politischen Bereich transferieren. Gerade im Bereich der Informationstechnologie wäre dies fatal, da sich abzeichnet, daß sie in allen Wirtschaftszweigen zur Schlüsseltechnologie avanciert und natürlich die Entwicklung trotz aller Schwächen immer noch am besten von den Experten eingeschätzt werden kann.

Siekmanns zeitlich differenzierte Betrachtung bezieht sich auf die Teilgebiete Sprachverstehen, Expertensysteme, Robotik, Deduktionssysteme und Computersehen. Für die Sparte Sprachverstehen wurde vorhergesagt, daß KI-Systeme binnen fünf Jahren - ausgehend vom Jahre 1984 - Befehlseingaben in geschriebener Sprache für eingeschränkte Bereiche problemlos verarbeiten können. Mittelfristig würde diese Fähigkeit durch gesprochene Ausgabe und das Verstehen bestimmter gesprochener Anweisungen erweitert. Langfristig sollten Maschinen komplexe Texte verstehen und ebenso komplexe Texte in natürlicher Sprache ausgeben können. Aus heutiger Perspektive läßt sich über die Vorhersagen des maschinellen Sprachverstehens folgendes Fazit ziehen: Die Prognosen über die Entwicklung natürlichsprachlicher

* Durch den Referenten freigegebene Wiedergabe seines Vortrages.

Systeme waren in Bezug auf die industrielle Umsetzung großenteils falsch, wenn auch nicht aufgrund einer prinzipiellen Überschätzung der technischen Machbarkeit, sondern weil die Extrapolationen wirtschaftliche Aspekte ausblendeten. Es wurde zu wenig berücksichtigt, daß technisch realisierbare Projekte einfach unökonomisch sein können. Das zeigt sich besonders am Beispiel der geschriebenen Befehlseingabe. Sie ist technisch durchaus realisierbar, aber deutlich zeitaufwendiger als etwa die Steuerung durch Menue und Maus, und somit für viele Anwendungen wirtschaftlich uninteressant. Die Schwelle zur Wirtschaftlichkeit natürlichsprachlicher Systeme dürfte mit der Beherrschung gesprochener Sprache als Ein- und Ausgabemedium gekoppelt sein. Letzteres hatte das Gutachten richtig als langfristige Perspektive eingestuft.

Ebenfalls anders als prognostiziert ist die Entwicklung bei Systemen zur automatischen Rohübersetzung verlaufen. Sie wurden erst für die Jahrtausendwende für möglich gehalten, weil davon ausgegangen wurde, daß ein zu übersetzender Text seitens der Maschine vollständig verstanden werden müsse. Es hat sich aber gezeigt, daß eine volles Verstehen des Textes ebensowenig notwendig ist wie eine genaue Übersetzung: Auch für Übersetzungen minderer Qualität gibt es ein Marktpotential, das automatische Rohübersetzungen bereits heute (also weit eher als vorausgesagt) zum technischen Standard gemacht hat. Eine weitere Fehleinschätzung attestiert Siekmann der Studie auch im mittelfristigen Bereich: Weder die Bürokommunikation noch die wissenschaftliche Kommunikation ist wie vorhergesagt integraler Bestandteil der Arbeitssituation geworden, sondern wird sich erst sehr viel langsamer umsetzen. Ähnliches gilt für Expertensysteme. Eine Mensch-Maschine-Symbiose im Bereich der Expertensystemtechnik zeichnet sich im Gegensatz zu den Prognosen des Gutachtens auch erst sehr viel langfristiger ab. Zwar werden in eng eingegrenzten wissenschaftlichen Gebieten mit beträchtlichem Erfolg Expertensysteme eingesetzt, jedoch völlig anders, als ursprünglich erwartet. Wie bei natürlichsprachlichen Systemen wurde die Wirtschaftlichkeit der KI-Komponenten deutlich überschätzt. In der Praxis hat sich gezeigt, daß isolierte KI-Systeme auf spezieller Hardware kaum durchsetzbar waren. Der Trend geht eindeutig zur Einbettung von KI-Komponenten in die Standardinformatik, und dort haben sie dann allerdings oftmals enormen Erfolg. Parallelen dazu zeigen sich auch auf dem Hardwaresektor. Die Zuwachsraten für die zu Beginn der achtziger Jahre auftauchenden Lisp-Maschinen wurden deutlich zu optimistisch eingeschätzt. Es wurde übersehen, daß die dramatische Leistungssteigerung der Standardhardware, die partiellen Vorteile von Lispmaschinen leicht kompensieren kann.

Fazit: Der Trend geht eindeutig zur Einbettung von KI-Systemen in die Standardinformatik, die technisch auf schnellen Standardmaschinen realisiert wird. Langfristig wurde bei Expertensystemen prognostiziert, daß sie ein integraler Bestandteil jeglicher fortgeschrittenen

Technologie werden, insbesondere im Maschinenbau; mehr noch, durch die massive Verbreitung von Expertensystemen würde langfristig ein neues Maschinenverständnis hervorgerufen. Diesen Wandel des Maschinenbildes hält Siekmann immer noch für durchaus realistisch (allerdings eventuell in einem anderen Zeithorizont). So geht die Industrie beispielsweise im Maschinenbau dazu über, Techniken der Standardinformatik zu integrieren und mit KI-Komponenten zu kombinieren. Mittlerweile sind diese Techniken unverzichtbar, und sie sind sicher nur die Spitze eines Eisberges.

Im Robotikbereich wurde vorhergesagt, daß fortgeschrittene Techniken der Informatik (beispielsweise spezielle Programmiersprachen für die Robotik), wenngleich ohne KI-Komponente, implementiert werden und erst langsam KI-Technologie zum Einsatz komme. Die Vorhersage im Robotikbereich war in gewisser Weise sehr konservativ gehalten, weil sich bereits zum Zeitpunkt der Gutachtenerstellung abzeichnete, daß der Einsatz von Robotik- und Expertensystemen im Vergleich zur menschlichen Arbeitskraft zu kostenintensiv ist. Die Praxis hatte gezeigt, daß in vielen Fällen die Verlagerung der Produktion in Billiglohnländer bzw. der Einsatz von Gastarbeitern dem massiven Einsatz von Robotiktechnologie vorgezogen wurde. Nicht vorhergesehen wurde, daß manche Länder - Japan vor allem - massiv auf robotikunterstützte Fertigung setzten, obwohl die Schwelle zur Wirtschaftlichkeit noch nicht erreicht war und sie sich damit dann einen heute kaum noch aufzuholenden Vorsprung in der Produktion erworben haben.

Weiter wurde ein Preisverfall und daraus resultierend eine massive Verbreitung von Robotiktechnologien über die rein industriellen Anwendungen hinaus vorhergesagt. Aber im Gegensatz zum starken Preisverfall bei Personalcomputern verzögert sich diese Entwicklung in der Robotik. Siekmann: "Es fehlt der Steven Jobs der Robotiktechnologie".

Langfristige Perspektiven im Robotikbereich: Wie vorhergesagt, zeichnet sich auch langfristig nicht die Entwicklung von anthropomorphen Robotern ab, und die industrielle Fertigung wird nicht von "Humunculi" übernommen werden. Die Gründe dafür liegen nicht in grundsätzlichen technischen Schwierigkeiten bei der Konstruktion anthropomorpher Systeme, sondern sind vor allem ökonomischer Art: Es ist weitaus billiger, eine Fabrik so umzubauen, daß automatische Fertigungsstraßen installiert werden können, als Roboter zu konstruieren, die sich in für manuelle Arbeitsabläufe ausgelegten Fabriken orientieren und Menschen ersetzen können. Daher zeichnet sich eine Entwicklung zu weitgehend automatisierter (nicht vollautomatischer) Fertigung mit dezentralisierten Teilsystemen und begrenztem Robotikeinsatz ab, die ihrerseits dezentral, aber auch von einem Zentralrechner überwacht werden. Für viele in-

dustrielle Bereiche ist eine vollautomatische Fertigung zwar technisch machbar, aber ökonomisch nicht sinnvoll.

Die Vorhersagen im Bereich der industriellen Umsetzung waren nicht immer richtig, vermochten aber in vielen Bereichen auf wichtige Trends aufmerksam zu machen. Grob falsch waren jedoch Zeitangaben, die die wirtschaftliche Durchsetzung bestimmter Teilgebiete der KI betrafen. Im allgemeinen ist ein S-Faktor von 3-5 anzunehmen. Der Anteil der KI an der gesamten wirtschaftlich genutzten Informationstechnologie ist wesentlich geringer, als 1984 prognostiziert wurde. Die KI selber hat kaum eine eigenständige wirtschaftliche Bedeutung, sondern sie ist in die Standardinformatik eingebettet, dort dann allerdings oftmals unverzichtbar und von entscheidendem Wettbewerbsvorteil.

Wissenschaftliche Aspekte

Ebenfalls ist die wissenschaftliche Entwicklung des gesamten Gebietes, wo sich neue, damals nicht absehbare Denkweisen etablieren konnten, nicht richtig gesehen worden. Die Studie ist entstanden im Tenor der rein symbolischen Repräsentation von Wissen in der Künstlichen Intelligenz. Es ist damals nicht erkannt worden, daß es wissenschaftliche Ansätze gibt, die die methodische Bedeutung einer rein symbolischen Repräsentation radikal in Frage stellen, zum Beispiel, die Theorie der neuronalen Netze. Die Gegenpositionen zur harten KI-These konvergieren derzeit in der Prämisse, daß die Beschaffenheit des Gehirns (also der Hardware) sehr wohl als bedeutsam für kognitive Prozesse anzusehen ist. Die Vorstellung, daß Denken an eine symbolische Repräsentation gekoppelt sein muß, ist eine aus der 2000 Jahre alten westlichen Wissenschaftstradition entstandene Vorstellung, die für den physischen Bereich, also das körperliche Sich-bewegen in der Welt beispielsweise nicht notwendig richtig ist. Es ist denkbar, daß kein physisches Zentrum gefunden wird, welches qua Repräsentation diese Bewegung steuert und koordiniert. Der Ansatz von R. Brooks, daß einfache und sicherlich nicht intelligente Teilsysteme miteinander kommunizieren und durch die raffinierte Kommunikation dieser Teilsysteme die Fähigkeit des Gesamtsystems ermöglichen, sich in der Umwelt zu bewegen, ist für motorische Intelligenz offensichtlich erfolgreich. Ähnliche Ansätze finden sich u.a. bei Maturana und Varela und Winograd/Flores, die den - auch für Naturwissenschaftler nachvollziehbaren - Gedanken propagieren, daß Intelligenz nicht nur in einem Wesen zu finden sei, das symbolisch repräsentieren kann, sondern in einem Wesen, das in der Welt "situiert" ist und durch kommunikative Prozesse mit ihr in Verbindung steht. Die Intelligenzleistung resultiert aus dem Zusammenspiel des Wesens mit der Welt. Dann wird auch denkbar, daß das kommunizierende Wesen in sich anders strukturiert ist, als es die harte KI annimmt, zum Beispiel nicht notwendig über eine abstrakte Repräsentation verfügt. Die span-

nende Frage in der gegenwärtigen KI-Forschung ist daher das Zusammenspiel solcher expliziten (symbolischen) Repräsentationen mit impliziten (subsymbolischen) Repräsentationsformen. Als Folge daraus müssen eventuell auch technische Fragen, wie die nach der Semantik, ganz anders gestellt werden, als es bisher der Fall ist. Zwar werde man die Konzepte der Tarski-Semantik weiter verwenden, aber es sei fraglich, ob dies allein die KI weiterführen könne. Die entscheidende Frage ist nicht die nach einer formalisierten Semantik, sondern nach deren Kontexttauglichkeit, d.h., nach der Bedeutung, die ein Satz oder eine Aktion für ein bestimmtes Individuum hat. Dieser Frage hat sich die Logik bisher kaum zugewandt.

Zusammenfassend läßt sich jedoch sagen, daß die Studie trotz dieser Mängel im großen und ganzen die Entwicklung erstaunlich richtig eingeschätzt hat.

Round-table-Gespräch über die "harte" KI-These: Der Mensch als Informationsverarbeitungssystem?

*Teilnehmer: Prof. Dr. R. Eckmiller
Prof. Dr. G. Görz
Prof. Dr. S. Krämer
Prof. Dr. G. Strube*

Moderation: Prof. Dr. I. Wachsmuth

Einleitung I. Wachsmuth

Zur Einleitung des Podiums zur "harten" KI-These werden vom Moderator folgende Fragen aufgeworfen:

Die "Integrationsfrage"

Integriert oder desintegriert die Leitvorstellung vom Informationsverarbeitungssystem das Verhältnis von Moralität und Kognition? Was wären Konsequenzen für die Verantwortbarkeit des Einsatzes künstlicher kognitiver Systeme?

Die "Komplexitätsfrage"

Penrose schlägt die Diskussion eines kritischen Grades an Komplexität vor, an dem Bewußtsein sich von allein einstellt. Für biologische Systeme scheint das durchaus plausibel. Inwieweit ist das Modell auf Maschinen übertragbar?

Die "Determiniertheitsfrage"

Könnte es sein, daß die Leitvorstellung des Informationsverarbeitungssystems nur auf dem Boden der klassischen Physik gültig und relevant ist? Hans Jonas macht z.B. darauf aufmerksam, daß neurologische Vorgänge quantensensibel, also nicht vorhersehbar sein könnten.

Wenn der Mensch ein informationsverarbeitendes System ist, ist er dann auch ein vollständig determiniertes System? Wenn nein, was ist er dann? Wenn ja, wo bleibt dann Raum für Ethik, Verantwortung und freien Willen?

Es folgen die Positionsbestimmungen der vier Podium-Teilnehmer:

Beitrag R. Eckmiller

Zu der zentralen Frage: "Ist der Mensch ein Informationsverarbeitungssystem?" wird ein fiktiver Indizien-Prozeß: Mensch A/Mensch B inszeniert: Mensch A behauptet: Ja, ich und wir alle sind Informationsverarbeitungssysteme. Mensch B behauptet: Nein, wir sind keine Informationsverarbeitungssysteme.

Situation: Es gibt nach allgemein von uns Menschen akzeptierter Meinung einerseits Pflanzen, Tiere, Menschen (Naturkreationen), andererseits Maschinen, Automaten, Computer (Menschkreationen).

Formale Aspekte der Prozeßführung:

1. Frage der Fachkompetenz von Mensch A und Mensch B
2. Frage der Befangenheit (Religion, Ideologie, Ängste)
3. Frage der Einsichtsfähigkeit (Kapazität der Kognition von A und B)
4. Frage der Interessenlage von A und B an diesem oder jenem Urteil
5. Technische Systeme als Mensch-Kreationen im Vergleich zu Biologischen Systemen als Natur-Kreationen (Dabei ist der Begriff Natur hier ein 'Platzhalter' für die uns Menschen unbekannte Ursache des Universums)

Methodik der Prozeßführung:

1. Feststellung gemeinsamer bzw. gegensätzlicher Positionen von A und B in:
 - a. Sprache
 - b. Weltbild, Philosophie
 - c. Naturwiss. Fachwissen (Physik, Biologie, Medizin)
 - d. Geisteswiss. Fachwissen (Philosophie, Psychologie, Soziologie)
 - e. Definition eines Menschen
 - f. Definition eines Informationsverarbeitungs-Systems
 - g. Sinn des Lebens, Sinn der Natur
 - h. Tabus

2. Festlegung der Methoden zur Indiziensammlung

- a. Naturwissenschaftliche Methoden
- b. Mathematische Methoden
- c. Geisteswissenschaftliche Methoden

3. Abschätzung der grundsätzlichen Beantwortbarkeit der Streitfrage

- a. Sind die Definitionen von Mensch bzw. Informationsverarbeitungssystem genau genug?
- b. Reichen die gegenwärtig verfügbaren Indizien aus?
- c. Ist es grundsätzlich denkbar, das eine Natur-Kreation (Mensch) seinerseits menschenähnliche Kreationen kreieren kann?
- d. Wie mental ähnlich müssen zwei Menschen A und B sein für ein gleiches Urteil?

Dieses fiktive Prozeßszenarium macht deutlich, daß der einzelne Mensch aufgrund seiner begrenzten kognitiven Fähigkeiten zu einem abschließenden Urteil nicht in der Lage ist. In dieser Situation bleibt dem einzelnen Menschen nur der Ausweg, sich für eine der möglichen Positionen als Arbeitshypothese zu entscheiden. Dabei ist daran zu erinnern, daß seit Descartes die Arbeitshypothese besteht, alle Lebewesen seien vollständig nach den Gesetzen der Physik beschreibbar (mit der meines Erachtens überzogen interpretierten Einschränkung für kleine Teilbereiche des Menschen). Nach dieser Arbeitshypothese, die bis heute nicht beweisbar ist, arbeiten Biologen, Mediziner und Biophysiker. Die notwendige Alternative wäre die ständige Beteiligung z.B. von Theologen an Biosubstrat-Forschungen.

Beitrag G. Görz

Vorbemerkung: Für den Diskurs in der Künstlichen Intelligenz und über die Künstliche Intelligenz ist ein extrem metaphorischer Sprachgebrauch kennzeichnend. Dies liegt zum einen an ihrem Gegenstand, ist darüberhinaus aber kennzeichnend für alle jungen Disziplinen. Schon die wörtliche Übersetzung von "artificial intelligence" durch "Künstliche Intelligenz" ist nicht unproblematisch, denn sowohl "artificial" als auch "intelligence" haben im Englischen nicht genau dieselbe Bedeutung wie "künstlich" und "intelligent" im Deutschen.

These 1: Es gibt keine stichhaltigen Argumente, die für die starke KI-These sprechen.

Die starke KI-These reduziert Kognition auf bloße Informationsverarbeitung. Es wird kaum bestritten, daß Kognition *auch* Informationsverarbeitung ist (das entspricht der schwachen

KI-These); die starke KI-These behauptet aber mehr, nämlich, daß Kognition nichts anderes als Informationsverarbeitung sei. Ein solcher Nachweis konnte aber bisher nicht erbracht werden. Die Behauptung, es sei *im Prinzip* der Fall, kann den Nachweis nicht ersetzen.

These 2: Die starke KI-These ist weder für die wissenschaftliche Grundlegung der KI noch für ihre technische Anwendung notwendig.

Die Grundlegendiskussion in der KI war von Anfang an stark durch das Problemlösungsparadigma bestimmt. Unbestreitbar hatte die "physikal symbol systems hypothesis" als Präzisierung des Informationsverarbeitungsparadigmas einen entscheidenden forschungsorientierenden Einfluß, der zur Ausformung der "symbolischen KI" führte. Ihre Verabsolutierung führt aber eher zu einer Einengung des Forschungsprogramms. Bei einer kritischen Betrachtung der obigen Argumentation von Herrn Strube kann ich nicht erkennen, an welchen Stellen von dieser Verabsolutierung faktisch Gebrauch gemacht wird - vielmehr ist die schwache KI-These hinreichend. Insofern kann ich Herrn Strube weitgehend zustimmen, ohne die starke KI-These als Prämisse annehmen zu müssen. In ihrer radikalen Formulierung impliziert die starke KI-These einen Substanzdualismus, der gerade auch der neueren Kognitionswissenschaft nicht haltbar erscheint. Es gibt eine Reihe signifikanter Hinweise auf eine enge Abhängigkeit der Kognition von der neuronalen Konnektivität. Zudem leistet die starke KI-These keinen Beitrag zu einer Erklärung des evolutionären Charakters der menschlichen Kognition.

These 3: Komplexität ist eine notwendige, aber nicht hinreichende Bedingung für Kognition.

Daß wir komplexen Systemen ein "Verhalten" zuschreiben, das wir mit anthropomorphen Termini charakterisieren - etwa, daß "Intelligenz" oder gar "Bewußtsein" emergieren würden - ist Ausdruck theoretischer Defizite. Solange nicht gezeigt werden kann, daß und wie intentionale Termini - auch "Intention" selbst - operationalisierbar sind, handelt es sich um nichts anderes als metaphorischen Sprachgebrauch. Zwar wird dadurch ein Programm benannt, aber in keiner Weise gelöst, denn erst die Operationalisierung würde die Adäquatheit psychologischer und sozialwissenschaftlicher Termini für die Beschreibung komplexer technischer Systeme aufweisen. Ich bezweifle die Möglichkeit eines durchgängig reduktionistischen Ansatzes und möchte demgegenüber in Anlehnung an den aktuellen kognitionswissenschaftlichen Diskurs für eine umfassende holistische Sichtweise der Kognition plädieren.

These 4: Intelligenz konstituiert sich in der sozialen Interaktion.

Ich will den Vorschlag aufgreifen, den Intelligenzbegriff so zu gebrauchen, daß wir ihn im sozialen Handlungszusammenhang verankern, also von "Intelligenz für uns" sprechen. Dann ist es auch unproblematisch, ihn auf die Interaktion von Mensch-Maschine-Systemen zu übertragen. Es besteht dann nicht die Notwendigkeit, einem technischen System, das uns als Medium bei Problemlösungen unterstützt, Intelligenz per se zuzuschreiben - die Intelligenz manifestiert sich in der Interaktion. Damit rückt zugleich auch die ethische Dimension in ein neues Licht: Relevant für das Menschenbild ist der Handlungszusammenhang, und in diesem Kontext ist das Verantwortungsproblem zu diskutieren.

Beitrag S. Krämer

1. Die Künstliche Intelligenz rückt in einen Brennpunkt des öffentlichen Interesses nicht als eine Ingenieurstechnik, sondern als eine Idee. Diese Idee besteht darin, daß "Intelligenz", "Denken" und "Geist" nicht länger mehr als anthropozentrische Kategorien zu behandeln seien, sondern im Prinzip auch einer Maschine zuzuschreiben sind. Die Künstliche Intelligenz-Forschung wird Bestandteil einer "kopernikanischen Wende" in unserer Auffassung vom Geist.

Es gibt eine Kluft zwischen der nahezu marginalen betriebswirtschaftlichen und sozialen Bedeutung von Techniken Künstlicher Intelligenz und der Emphase, mit der Verfechter und Kritiker Künstlicher Intelligenz Kongresse, Feuilletons und akademische Seminare zu beschäftigen wissen. Diese Kluft wurzelt darin, daß die Künstliche Intelligenz die erste Technik ist, deren ideelle Folgeerscheinungen ihre faktischen Folgelasten überwiegen. Künstliche Intelligenz wirkt nicht als eine Ingenieurstechnik, sondern als eine Idee. Genauer: Als ein Modell für die Erklärung von Geist. Um die Sprengkraft dieses Modells zu verstehen, lohnt ein Blick in die Geschichte der Theorien vom Geist. Im antiken Griechenland galt der Geist als ein kosmologisches Ordnungsprinzip, welches im christlich-romanischen Sprachraum umgedeutet wurde zur göttlichen Schöpfungspotenz und schließlich trinitätstheologisch aufgewertet wird zum Heiligen Geist. Erst die Neuzeit verortet den Geist im denkenden Individuum: In diesem Sinne gebraucht René Descartes "Geist" und "Bewußtsein" synonym. Nicht Gott, vielmehr der Mensch ist fortan privilegiert durch seinen Zugang zum Geist. Seit der Mitte dieses Jahrhunderts wird Geist nun thematisiert in den Termini von Informationsverarbeitung, deren Vorbild wiederum gewonnen ist an der Datenverarbeitung durch den Computer. Zeugnis dieser Vorbildfunktion ist die jüngst entstandene Kognitionswissenschaft, die das Denken als ein Rechnen über

mentalen Repräsentationen definiert. Die Karriere des Informationsverarbeitungsmodells in nahezu allen Wissenschaften, die das Denken untersuchen, signalisiert, daß der anthropozentrische Geistbegriff der Aufklärung offensichtlich abgelöst wird durch einen nichtanthropozentrischen, durch einen maschinenzentrierten Begriff des Geistes. Wir sehen also: Erst Gott, dann der Mensch und schließlich die Maschine werden zu den sich ablösenden Leitbildern, an denen die Antike, die Neuzeit und die Moderne ihren Entwurf vom Geist jeweils ausrichten. Nachdem Kopernikus den Menschen aus dem Zentrum des Universums rückte, nachdem Darwin ihm das Vorrecht privilegierter Erschaffung und Freud ihm die Auszeichnung durch die handlungsleitende Instanz des Bewußtseins nahm, scheint der Mensch nun eines seiner letzten Privilegien zu verlieren: Daß nämlich der Geist eine Auszeichnung der menschlichen und nur der menschlichen Spezies sei. Die Künstliche Intelligenz wird zum Vorreiter dieser Desillusionierung. Vielleicht erklärt dies die ideologieverdächtige Leidenschaft, die eine praktisch so wenig genutzte Technik bei ihren Gegnern wie bei den Befürwortern provoziert.

2. Die Aktivität externer Manipulation von Symbolen kann zur technischen Verstärkung von Intelligenz eingesetzt werden. Weite Bereiche unseres Denkens beruhen auf einer symbiotischen Kopplung von Intellekt und Symboloperation. Die Technik symbolischer Maschinen, die wir vom schriftlichen Rechnen kennen, bildet eine der wichtigsten Gestalten symbolischer Geistestechniken. Mit ihr werden komplizierte geistige Tätigkeiten zurückführbar auf das mechanische Operieren mit externen Symbolen. Künstliche Intelligenz beginnt nicht mit dem Computer, sondern mit der intellektuellen Nutzung symbolischer Maschinen.

Der Mensch ist nicht nur, wie Aristoteles meinte, ein vernunftbegabtes Lebewesen, sondern mit den Worten des Philosophen Ernst Cassirers ein symbolgebrauchendes Lebewesen. Auf Symbolisierung sind wir nicht nur angewiesen, wenn wir kommunizieren, sondern auch dann, wenn wir denken: 7777 mal 8888 im Kopf auszurechnen gemahnte eher an eine akrobatische Kunst. Daß wir Rechenaufgaben mit künstlichen Hilfsmitteln lösen müssen, seien das nun Rechenbrett, Papier und Bleistift oder Taschenrechner, zeigt, wie sehr Bereiche unseres Denkens immer schon gestützt werden durch Geistestechniken der Symbolmanipulation. Eine herausragende Rolle unter diesen Geistestechniken spielen symbolische Maschinen. Unter einer symbolischen Maschine sei ein kalküliertes Zeichensystem verstanden, auf welches ein Bereich von Gegenständen so eindeutig abgebildet werden kann, daß Probleme, die diese Gegenstände betreffen, durch rein mechanische Manipulationen innerhalb dieses Zeichensystems lösbar werden. Beispiel dafür ist unser arithmetischer Kalkül. Wo symbolische Maschinen zum Einsatz kommen, kann die syntak-

tische Konsistenz der Symboloperation die Garantie übernehmen für deren semantische Kohärenz. Wahrheit wird so zurückführbar auf Richtigkeit. Wo das der Fall ist, können geistige Tätigkeiten auf geistlose Weise durchgeführt werden. Geist wird ohne Bewußtsein realisierbar. Mit der Durchsetzung des schriftlichen Rechnens im dezimalen Positionssystem im Europa des 14. und 15. Jahrhundert, wird der Gebrauch einer symbolischen Maschine zu einer Kulturtechnik. Hierbei bleibt die Aktivität der Symbolverarbeitung, durch welche die Leistungen des Intellekts gesteigert werden, noch manuell. Mit dem Computer entsteht die Möglichkeit, die externe Symbolmanipulation zu automatisieren. Eine ungeahnte Steigerung von Vorgängen der Symboltransformation ist die Folge. Im Sinne symbolischer Intelligenzverstärker gibt es Künstliche Intelligenz also, seit es symbolische Maschinen gibt. Entgegen der Suggestion unseres Sprachgebrauches daß, wo immer etwas künstlich hervorgerufen werde, es ein natürliches Pendant geben müsse, gibt es nicht einfach ein natürliches, ein naturbelassenes Denken. Unsere intellektuellen Fähigkeiten sind immer schon mediatisiert durch die symbolischen Leistungen der Sprache, der Schrift und der Bilder.

3. Das Modell vom Denken als Informationsverarbeitung überträgt Max Webers Typus des zweckrationalen Verhaltens auf den Bereich geistiger Tätigkeiten. Dieses Modell erfaßt nur die zieladaptierenden, nicht aber die wertsetzenden Leistungen des Intellekts. Die Fähigkeit, Werte nach Maßgabe individueller Präferenz zu setzen, verkörpert eine Dimension menschlicher Intellektualität, die mit dem Informationsverarbeitungsansatz nicht zu modellieren ist.

Die symbolorientierte wie die konnektionistische Künstliche Intelligenz sind der Überzeugung verpflichtet, daß die Intelligenz eines Systems in der Fähigkeit bestehe, seine Ziele angesichts von Schwierigkeiten und Verwicklungen in der Aufgabenumgebung zu realisieren. Die am John-von-Neumann-Computer und die an Neuronalen Netzen orientierte Künstliche Intelligenz sind nur unterschiedliche Strategien zur Realisierung adaptiver Informationsverarbeitung. Dies nimmt ihrer Konfrontation die Schärfe. Nun ist zieladaptives Verhalten zwar eine notwendige, nicht jedoch hinreichende Bedingung menschlicher Intelligenzleistung. Neben der Fähigkeit zur Ziel- und Zwecksetzung haben wir auch die Fähigkeit, Werte zu setzen, "Wert" hier nicht verstanden als eine entscheidungstheoretisch modellierbare Folgen-Nutzen-Evaluierung, sondern verstanden als eine Setzung, die ihrerseits nicht wieder ableitbar ist aus anderen Werten, sondern die eine genuine Entscheidung eines Individuums verlangt angesichts von Präferenzen, die in seiner Selbstinterpretation wurzeln. Kaufen wir ein Auto, so könnte wir uns entscheiden mit Hilfe eines Kalküls, welcher aus der Relation zwischen Ansprüchen und Leistungen den

optimalen Autotyp errechnet. Nur: So entscheiden wir gemeinhin nicht. Vielmehr werden wir gelenkt durch Präferenzen, die zu tun haben mit unserer Selbstprojektion, mit unserem Lebensstil und individuellen Präferenzen. Im Unterschied zu zweckrationalen Entscheidungen kann über die in Selbstinterpretation gründenden Entscheidungen nicht noch einmal im Sinne von "richtig" oder "falsch" entschieden werden. Das meint Carl Schmitt, wenn er sagt, "die Entscheidung ist, normativ betrachtet, aus dem Nichts geboren". Solche durch Wertsetzung gesteuerten Entscheidungen sind originäre Akte von Willensbildung, die weite Bereiche unseres alltäglichen Denkens und Tuns durchdringen. Sie sind mit dem Ansatz der Informationsverarbeitung nicht zu modellieren. Wo immer wir einem Akteur Willensfreiheit zubilligen, behandeln wir ihn als Person. Die wesentliche Auszeichnung des Menschen, kraft derer er zur Person wird, liegt also nicht im Besitz eines Intellekts, sondern eines Willens. Das Zusammenspiel beider ist für geistige Leistungen konstitutiv.

4. Technische Apparate sind immer auch Projektionen unseres Selbst- und Weltverständnisses. Wo der Computer zum Modell wird für den Geist, ist ein kognitivistisches Menschenbild impliziert. "Kognitivistisch" heißt: Die Leistungen des Intellekts werden abgespalten von der Urteilskraft. Das Menschenbild Künstlicher Intelligenz beruht auf der Trennung von intellektueller und moralischer Kompetenz, von Intellekt und Wille.

Schon die rationalistischen Philosophen Descartes und Leibniz versuchten, die Intelligenzverstärkung mittels symbolischer Maschinen für den gesamten Bereich des wissenschaftlichen Erkennens fruchtbar werden zu lassen. Doch diese Erkenntnisteknik galt ihnen gerade nicht als Vorbild für eine Theorie des Geistes. Für die Philosophen der Aufklärung stand außer Frage, daß Geist gebunden sei an die Eigenschaft des Menschen, nicht nur über einen Intellekt zu verfügen, sondern eine Person zu sein. Descartes unterscheidet zwei Vermögen des Geistes, den Verstand und den Willen. Der Verstand ist die intellektuelle, der Wille aber die moralische Instanz, und das heißt für Descartes "Sitz" der Urteilskraft. Denken geht hervor aus dem Zusammenspiel beider: Der Verstand produziert die Ideen; der Wille beurteilt diese als wahr oder falsch. Und John Locke unterscheidet zwischen "man" und "person". Unter einer Person versteht er ein Wesen, das über persönliche Identität verfügt und also rechenschaftsfähig und -pflichtig ist für sein Tun. Über Geist verfügen nach John Locke nun ausschließlich Personen, so daß also - wo Geist gegeben ist - auch Verantwortung zu nehmen sei. Zurück nun zur Künstlichen Intelligenz. Wo immer eine geistige Tätigkeit programmiert wird, ist diese in einer Weise umstrukturiert, daß ihre Ausführung an Urteilskraft nicht länger gebunden ist. Dies gelingt, wenn die Regeln, die die geistige Tätigkeit ursprünglich leiten, in Gestalt von Algorithmen vollständig expliziert werden. Algorithmen aber bilden keine Unterklasse von Regeln, sondern haben

den Status, eine Regel zu sein, gerade verloren. Denn die Fähigkeit, einer Regel zu folgen - dies ist Wittgensteins unhintergehbare Einsicht - setzt die Fähigkeit voraus, Geltungsbedingungen anzuerkennen und gegebenenfalls abzuerkennen; also Regeln auch brechen und sie im Handeln modifizieren zu können. Kurz: Regelfolgen sind gebunden an die Fähigkeit zur Urteilskraft. Solche Fähigkeit aber entsteht durch Teilnahme an einer öffentlichen Lebensform. Hier ist der Ort, wo die Kompetenz, auf Geltung von etwas zu reagieren, also die Unterscheidung des Faktischen und des Normativen entsteht. Wo der Computer zum Leitbild wird für das, was Geist sei, ist die Einheit von kognitiver und moralischer Kompetenz, von Intellekt und Wille aufgespalten. Erkennen soll möglich werden ohne anzuerkennen; Wissen soll zu gewinnen sein, ohne daß ein Gewissen gegeben ist. Leitbilder sind nicht danach zu beurteilen, ob sie richtig oder falsch sind, sondern danach, ob sie orientieren oder desorientieren. Wenn viele Probleme, vor denen wir gegenwärtig stehen, nicht durch ein Mehr an effizienter Information, sondern durch ein Mehr an Urteilskraft zu lösen wären, dann allerdings wirkt das kognitivistisch verkürzte Menschenbild desorientierend. Das Problem mit der Künstlichen Intelligenz liegt alleine darin, daß der Mensch beginnen könnte sich so zu verstehen, wie ihn die Künstliche Intelligenz modelliert.

Beitrag G. Strube

Voraussetzungen und Konsequenzen maschineller Intelligenz

I. Die Grundthese der KI

Die *physical symbol systems hypothesis*, jene Annahme von Newell & Simon (1976), die besagt, die Fähigkeit zur Symbolbildung und Symboltransformation sei eine notwendige und sogar hinreichende Bedingung für ein intelligentes System, trifft prinzipiell zu. (Dies sagend, vertrete ich die These der "harten" KI.) Hinsichtlich der Kognition unterscheidet der Mensch sich demnach nicht prinzipiell von anderen, zum Beispiel maschinellen intelligenten Systemen. Dies bedarf der Detaillierung, denn "prinzipiell" bezieht sich hier darauf, daß die *physical symbol systems hypothesis* ja nicht auf irgendetwas Beliebiges, was wir intelligent nennen könnten, sich bezieht, sondern einem Verständnis von künstlicher Intelligenz anhängt, wie es für die Gründerjahre dieser Disziplin typisch war, wo nämlich mit Simons Worten künstliche Intelligenz sich mit Problemen befaßt, zu deren Lösung beim Menschen typischerweise Intelligenz unterstellt wird.

Für den Begriff der "Künstlichen Intelligenz" gilt ohnehin, daß der Intelligenzbegriff, wie er beim Menschen angewendet wird, eigentlich in die Irre führt und gar nicht dasselbe meint wie

"künstliche Intelligenz", denn wir verwenden ihn zum Vergleich *zwischen* Menschen. So wie wir verschieden groß sind, so sind wir unterschiedlich intelligent, und vorausgesetzt ist, daß auch der - pardon - Dümme von uns in der Lage ist, Leistungen zu erbringen, zu deren Voraussetzungen auf technischem Gebiet die künstliche Intelligenz noch lange nicht gelangt ist. Künstliche Intelligenz hat ja gerade damit zu tun, jene Fähigkeiten im technisch-maschinellen Bereich zu realisieren, die bei uns noch lange nicht unter den Intelligenzbegriff fallen, plus einiger solcher, die dies dann tun. Und genau auf den letzteren Bereich, auf Denken, Gedächtnis, Sprache also, ist die *physical symbol system hypothesis* bezogen, in diesem oder für diesen Bereich ist sie entwickelt worden, und für diesen Bereich, behaupte ich, hat sie auch Gültigkeit. Unzweifelhaft gibt es Bereiche, wo wir ohne Symbolmanipulation auch in unserem Denken nicht auskommen.

2. Komplexe Architektur

Insofern also meine ich, daß die *physical symbol systems hypothesis* für den Kernbereich der Kognition gilt. Nun umfaßt ein intelligentes System, zumal ein *autonomes* intelligentes System, natürlich weitere Komponenten, die im wesentlichen den Charakter von Eingabe-Ausgabe-Komponenten haben, also Sensorik, Motorik, und all das, was an sensorischer Verarbeitung und motorischer Steuerung nötig ist. Zwar ist auch ein autonomes Subsystem wenigstens zum Teil der Kontrolle der zentralen kognitiven Verarbeitung unterworfen, aber ein autonomes intelligentes System völlig ohne solche wenigstens zum Teil selbständigen Komponenten wäre nicht einmal vorstellbar. Insofern müssen wir von diesem Kernbereich graduell nach außen gehen. Dabei wird immer fraglicher, was noch intelligent genannt werden soll oder was nur geschickte biologische Informationsverarbeitung ist, so ähnlich, wie sich die Frage stellt, ob die Steuerung ihres Fluges, die die Stubenfliege mit relativ wenigen Neuronen zustandebringt, eigentlich intelligent zu nennen ist oder nicht. Hier komme ich auf das Postulat von Newell & Simon zurück, es sei schon eine hinreichende Bedingung für ein intelligentes System, Symbole manipulieren zu können. Notwendig erscheint mir außerdem (hier stimme ich ausnahmsweise mit Penrose, 1989, überein), daß ein Mindestmaß an *Komplexität* gegeben ist, wobei ich sofort eingestehe, daß es zur gegenwärtigen Zeit niemanden gibt, der dieses Mindestmaß an nötiger Komplexität zu bestimmen in der Lage wäre. Also: Thermostaten denken nicht. Im Gegensatz zu Penrose bin ich aber der Ansicht, daß Komplexität allein weder Intelligenz noch gar Bewußtsein garantiert. Es kommt nicht einfach auf Komplexität an, sondern es kommt auf eine bestimmte *Architektur* an.

Ich will am Beispiel Bewußtsein versuchen, dies etwas zu präzisieren. Bewußtsein ist ja nicht etwas Funktionsloses. (Dies im Gegensatz etwa zu einer Position, wie sie der Marburger Philo-

soph Bieri vertritt, der davon ausgeht, die Kognition funktioniere ja ganz genauso auch ohne Bewußtsein, und von daher zu der Frage kommt, was denn dieses rätselhafte Bewußtsein soll.) Ich denke, Kognition würde ohne Bewußtsein anders ablaufen. Denn die Funktion des Bewußtseins ist zweierlei: einmal *monitoring*, d.h. Überwachung und Steuerung, Reflexion auch über das, was gerade abläuft (eine Komponente also, die in allen Systemen einer bestimmten Mindestkomplexität notwendig ist, um ein effizientes Arbeiten zu ermöglichen). Zum zweiten aber besteht die spezielle Funktion gerade auch des bewußten Überlegens darin, *Probehandeln* zu ermöglichen. Probehandeln muß aber, um in der Interaktion mit der Umwelt, in der jedes autonome intelligente System steht, erfolgreich zu sein, dieselben Kategorien verwenden, die erfolgreich sind für das konkrete Handeln. Und - so jedenfalls sieht es Wolfgang Prinz (1992) - dies ist der Grund dafür, daß wir nicht irgendwelche internen Vorgänge, sondern gerade *Objekte und Ereignisse* als die wesentlichen Kategorien unserer mentalen Repräsentationen haben, und daß wir sie zugleich nach außen verlegen, so daß wir die Objekte da draußen, wo wir sie anfassen können, sehen, und nicht etwa da in unserem Kopf, wo sie repräsentiert sind. Solchermaßen betrachtet, ist also die Genese von Bewußtsein zwar an Komplexität gebunden, aber folgt nicht einfach aus ihr.

3. Umwelteinbettung

Ein weiterer Punkt, den ich für wichtig halte und der auch von Newell (1980) schon betont worden ist, obwohl dies viele seiner Kritiker unterschlagen, besagt, daß ein kognitives System in eine Umwelt eingebettet ist. Newell nennt dies das *symbol grounding problem*. Die mentalen Repräsentationen müssen etwas *bedeuten*. Searle, der um die gleiche Zeit die KI fundamental kritisiert hat (Searle, 1980), stellt genau das in Frage anhand damals schon existierender KI-Programme, wie der derjenigen von Schrank. Er kritisiert, daß Intentionalität, im allgemeinen Sinn von Weltbezug verstanden, in solchen Maschinen nicht existiere. Dies ist m.E. nicht notwendig der Fall. Es ist nur so, daß das übliche KI-Programm, das auf einem normalen Tischrechner läuft, absolut defizitär in seinem Weltbezug ist, der schlicht darin besteht, Eingaben von einer Tastatur entgegenzunehmen und irgendetwas mit dem Bildschirm auszugeben. Ein autonomes System steht in der Notwendigkeit einer wirklichen Umwelteinbettung und ist darauf angewiesen, direkt über Sensorik und Motorik mit der Umwelt interagieren zu können. Über Sensorik und Motorik schließt sich auch die Schleife zwischen System und Umwelt, so daß wir hier also von einer Koppelung sprechen können, obwohl ich das nicht in dem restriktiven Sinn verstanden wissen will wie Maturana und seine Anhänger (z.B. Maturana, 1985), die wiederum in Abrede stellen, daß es sich hier überhaupt um Informationsverarbeitung handelt.

4. Konsequenzen für das Menschenbild

Abschließend will ich noch kurz auf das Menschenbild eingehen, und zwar dahingehend, daß Kognition oder gerade diese höheren kognitiven Prozesse immer dazu gedient haben, das Spezifikum des Menschen hervorzuheben. Der Mensch kann eben denken und sprechen, die Tiere können das nicht, also ist die Abgrenzung zur übrigen biologischen Welt vornehmlich durch Rekurs auf diese Fähigkeiten erfolgt. Nun sehen wir uns auf einmal in dem Dilemma, daß das, wodurch wir uns bisher vom Rest der Schöpfung abgesetzt haben, plötzlich dasselbe ist, was sogar Maschinen haben. Dagegen wiederum kann eine Abgrenzung erfolgen, indem man neuerlich das Biologische betont: die Leiblichkeit des Menschen, die Emotionalität, kurz, alle die Dinge, die im maschinellen Bereich nicht existieren. Hier wäre noch ein weiteres zu sagen: Künstliche Intelligenz im Wortsinne, wirkliche Kognition realisiert auf einer Maschine, wird mit Sicherheit eine *unmenschliche* Kognition sein, denn unsere menschliche Kognition interagiert ja gerade mit unserer biologischen Existenz und Leiblichkeit und vielleicht noch stärker - und dafür ausschlaggebend - mit unserer ganzen Erziehung, all dem, was man unter dem Stichwort Enkulturation bezeichnet. Dies alles dient als übergreifende Klammer zwischen Kognition und Enkulturation. Normen werden nämlich diskutiert in einer Gesellschaft und kulturell weitergegeben, d.h. also, die Trennung von Kognition auf der einen und biologischen wie sozialen und kulturellen Faktoren auf der anderen Seite wird aufgehoben durch eine Integration, die in der Person des Menschen stattfindet, qua dessen Sozialisation und Enkulturation. Dies können wir von Systemen der künstlichen Intelligenz nicht erwarten. Wir sollten uns bewußt sein, daß intelligente Computer andere Arten sind, nicht menschenähnlich, wenn auch intelligent.

Literatur:

- Bieri, P. (1992, Oktober). Was macht Bewußtsein zu einem Rätsel? *Spektrum der Wissenschaft*, pp. 48-56.
- Maturana, H. R. (1985). *Erkennen: Die Organisation und Verkörperung von Wirklichkeit* (2. Aufl.). Braunschweig: Vieweg.
- Newell, A. (1980). Physical Symbol Systems. *Cognitive Science*, 4, 135-183.
- Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*, 19, 113-126.
- Penrose, R. (1989). *The emperor's new mind*. Oxford: Oxford University Press.
- Prinz, W. (1992). *Prinzipien des Sehens* (Paper 6/1992). München: Max-Planck-Institut für Psychologische Forschung.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417-457.

Die im Anschluß daran durchgeführte Diskussion ist im Diskurs-Protokoll zur Technikfolgenabschätzung der Informationstechnik II - 3 veröffentlicht (Hrsg.: VDI/VDE-Technologiezentrum Informationstechnik GmbH, Dezember 1993, S. 14-22).

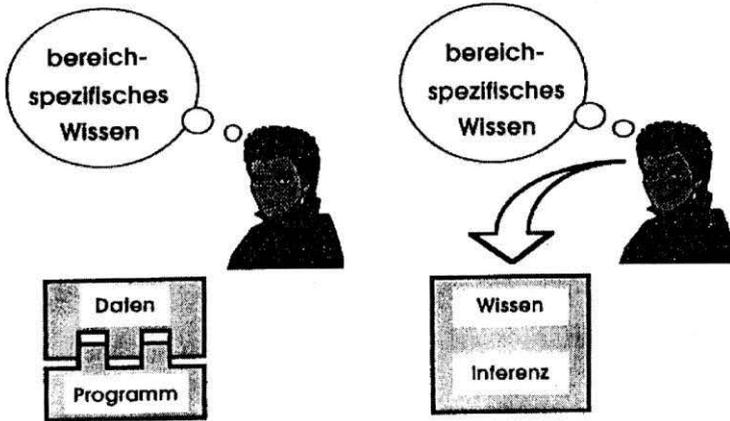
Künstliche Intelligenz als Technik - Hoffnungen, Erwartungen und heutiger Stand

Prof. Dr. G. Barth

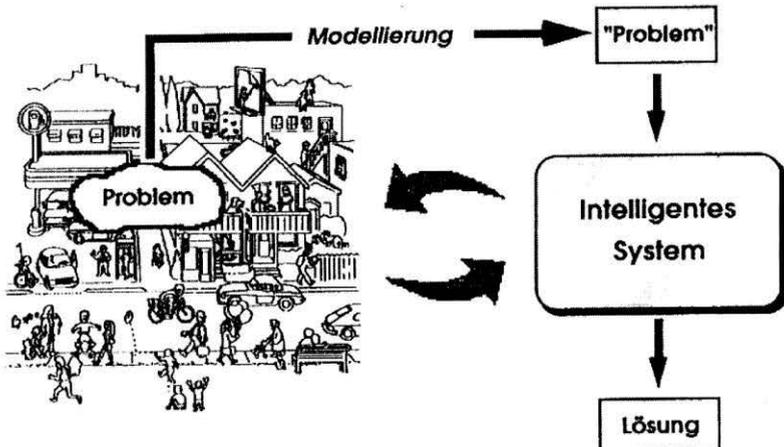
Es gibt wahrscheinlich so viele Definitionen der *Künstlichen Intelligenz* wie Textbücher. In diesem Aufsatz liegt allen Ansichten und Äußerungen die Vorstellung zugrunde, daß KI zu verstehen ist als die maschinelle Verarbeitung von Wissen. Computer sind uns Menschen längst überlegen in der Speicherung und dem Wiederfinden von Daten sowie deren Verknüpfung durch arithmetische, logische oder ähnliche Operationen. Bevor der Frage nachgegangen wird, inwieweit Maschinen diese Fähigkeiten im Umgang mit Wissen besitzen, soll zunächst der Versuch einer Abgrenzung zwischen Daten und Wissen unternommen werden. Daran schließt sich direkt die Frage nach den grundsätzlichen Unterschieden in ihrer Verarbeitung durch Computer an.

Wissen ist die Fähigkeit, Daten in einem gegebenen Kontext zielgerichtet interpretieren zu können. Ein Arzt etwa zieht aufgrund seines medizinischen Wissens die von ihm festgestellten Daten - wie Körpertemperatur, Puls, Herztöne - zur Erstellung einer Diagnose und einer Therapie heran. Ein Laie kann mit diesen Daten fast nichts anfangen, er kann sie höchstens nach ihm vorgegebenen Regeln manipulieren, sie z.B. kopieren oder in eine Tabelle eintragen. Hierin wird der Unterschied zwischen der Verarbeitung von Wissen und Daten recht deutlich. Natürlich benötigt der Entwickler eines datenverarbeitenden Systems eine ganze Menge an Fachwissen. Dieses wird allerdings nicht zum Bestandteil seines Produkts. Gerade andersherum verhält es sich bei wissensverarbeitenden Systemen, wo das Wissen in einer geeigneten formalen Darstellung zur wesentlichen Komponente des Computersystems wird.

Die Architektur wissensbasierter Systeme leitet sich unmittelbar aus dem alltäglichen Zusammenspiel zwischen Fachexperten und Laien ab. Letztere wenden sich mit ihren Problemen, die sie im wesentlichen mittels Daten beschreiben können, an Personen, welche diese Daten richtig interpretieren können. Die dabei ablaufenden Denkprozesse werden in Computersystemen durch geeignete Inferenzmechanismen simuliert. In einem Datenspeicher werden die das ursprüngliche Problem beschreibenden Werte sowie später die jeweils ermittelten Zwischenresultate abgelegt.



Es wird häufig an wissensbasierte Systeme die Forderung gestellt, Probleme auf genau die gleiche Art und Weise wie Menschen lösen zu können. Dies ist eine viel zu strenge Sicht der Dinge. Um uns herum gibt es eine Reihe von Beispielen, wo mit Hilfe von technischen Geräten schwierige Probleme vollkommen andersartig als von natürlichen Lebewesen gelöst werden. Man denke etwa an die grundsätzlichen Unterschiede zwischen dem Vogelflug und der Antriebstechnik moderner Flugzeuge. Warum also sollten wissensbasierte Systeme die Denkvorgänge von Menschen exakt nachbilden, um dadurch so schwierige Aufgaben wie die Erstellung von Diagnosen und Prognosen oder die Konfiguration komplexer Gebilde erledigen zu können?



Der folgende fiktive Dialog zwischen einem Menschen und einem Computer soll einige der Schwächen und Stärken heutiger wissensbasierter Systeme darlegen. Der Deutlichkeit halber sei gesagt, daß die Bemerkungen des Computers in Kursivschrift gedruckt sind...

Ich habe Allgemeinwissen

Ich habe fast kein Allgemeinwissen, jedoch eine Menge an Fachwissen.

Ich kann meine Kompetenz normalerweise von vornherein einschätzen.

Ich erfahre die Grenzen meiner Kompetenz immer erst dann, wenn ich Probleme anpacke.

Ich kenne die Kompetenz meiner Partner und kann mit ihnen flexibel kooperieren.

Ich kann die Kompetenz meiner Partner nur schwer einschätzen und bevorzuge feste Regeln der Zusammenarbeit.

Ich bin lernfähig und kann dabei verallgemeinern und spezialisieren.

Ich tue mich beim Lernen noch immer etwas schwer.

Ich kann Ausnahmefälle fast immer erkennen und meine Überlegungen entsprechend anpassen.

Ich kann Ausnahmefälle nur sehr schwer feststellen.

Ich bin nicht immer gleich gut in Form.

Ich kenne keine Schwankungen in meiner Leistungsfähigkeit.

Ich vergesse schon mal was.

Ich habe noch nie etwas vergessen.

Ich kann immer nur an einem Ort gleichzeitig sein.

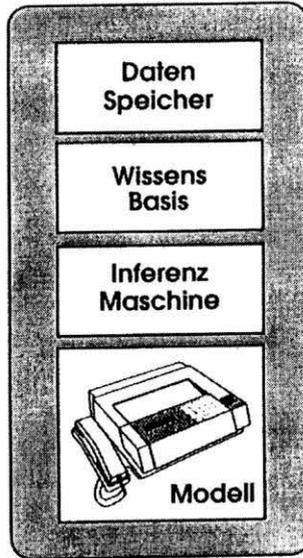
Mich kann man beliebig oft vervielfältigen.

Ich fühle mich an manchen gefährlichen Orten nicht wohl.

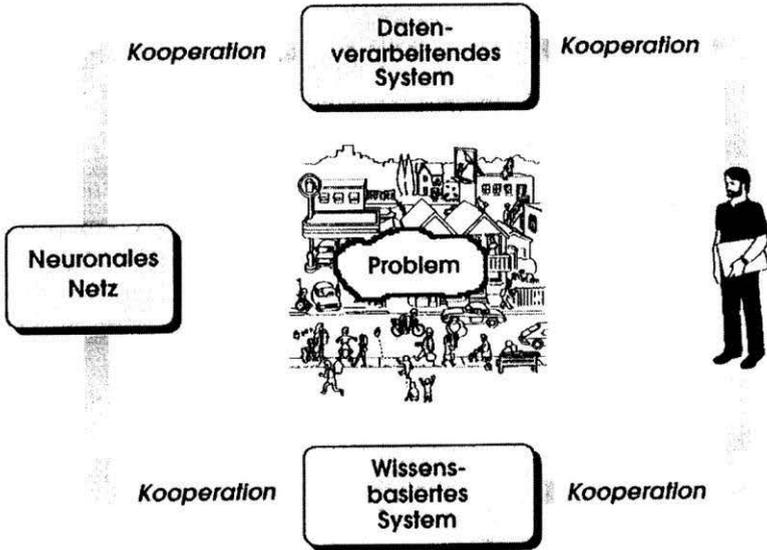
Mich kann man fast überall hinschicken.

Eine der größten Herausforderungen hinsichtlich der Konstruktion wissensbasierter Systeme ist es, sie mit leistungsfähigen Mechanismen für das Lernen neuer Erkenntnisse auszustatten. Es gibt eine Vielzahl von sehr interessanten Ansätzen, wobei in erster Linie neuronale Netze zu nennen sind. Trotz mancher beeindruckender Leistungen sind sie allerdings noch weit hinter den Möglichkeiten von Menschen zurück und werden dies wohl auch immer bleiben. Es ist einfach faszinierend, welche phantastischen Fähigkeiten das menschliche Gehirn beim Erfassen und Einprägen neuartiger Konzepte besitzt. Wir sind in der Lage, sowohl aus Fehlern als auch aus positiven Erlebnissen zu lernen, wir können verallgemeinern und spezialisieren, und wir erkennen scheinbar mühelos Analogien. Gemessen an diesen Kriterien kommen wissensbasierte Systeme bisher nicht einmal über das Stadium von kleinen Kindern hinaus.

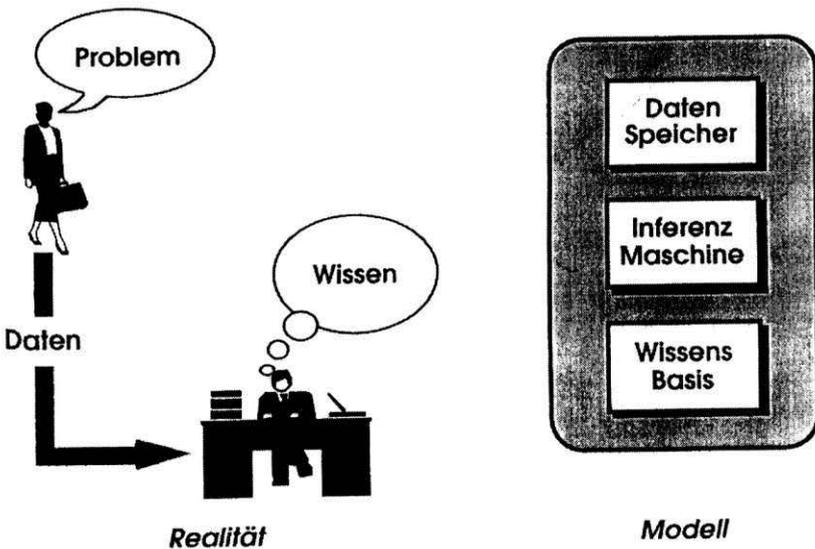
Heutige Systeme vertrauen normalerweise blind ihrer Wissensbasis und dem darauf agierenden Inferenzmechanismus. Sie besitzen kaum eine Möglichkeit, die Plausibilität oder tatsächliche Richtigkeit der einmal von ihnen gezogenen Schlußfolgerungen zu überprüfen. Rein wissensbasierte Systeme können auch nicht ihre eigene Kompetenz a priori überprüfen, vielmehr werden sie im Normalfall erst während der Bearbeitung einer Aufgabe an ihre Grenzen stoßen und aufgeben müssen. Sehr viel kritischer sind natürlich diejenigen Fälle, bei denen tatsächlich eine "Lösung" gefunden wird, diese allerdings in Wahrheit gar keine solche ist. Handelt es sich dabei um eine Anwendung zur Beratung eines menschlichen Experten, so wird dieser hoffentlich größeren Schaden verhindern können. Kritisch wird die Angelegenheit allerdings dann, wenn das wissensbasierte System als integrierte Komponente eines größeren Gebildes verwendet wird und dessen Gesamtfunktion beeinflusst. Um solchen Gefahren entgegenzuwirken, werden mehr und mehr Systeme neben einer Wissensbasis und einem darauf wirkenden Inferenzmechanismus mit einem Modell der Zusammenhänge und Vorgänge des betreffenden Anwendungsbereiches ausgestattet. Es gibt noch keinerlei festgefügte Formen solcher Modelle, typische Beispiele sind etwa Gleichungen oder Beschränkungsrelationen ("constraints"). Solche Randbedingungen können dann sowohl beim Erwerb von Wissen als auch bei der Herleitung von Folgerungen zu Kontrollzwecken herangezogen werden. Natürlich wirft dieser Gedanke die zusätzliche Schwierigkeit nach der Bereitstellung geeigneter Modelle auf. Dieses Problem ist eine der wichtigsten Fragestellungen für zukünftige Forschungsaktivitäten.



Die Zukunft wissensbasierter Systeme wird ganz entscheidend davon abhängen, inwieweit sie mit anderen Computersystemen zusammengefügt werden können. Die Komplexität wirtschaftsrelevanter Probleme macht die Integration verschiedener Ansätze zwingend notwendig. Weder regelbasierte Expertensysteme noch neuronale Netze oder Fuzzy-Kontroller sind in der Lage, alle Aufgaben gleichermaßen gut zu bearbeiten. Das Hauptaugenmerk weiterer Forschungsanstrengungen darf sich deshalb nicht auf die isolierte Betrachtung einzelner Ansätze und deren dogmatische Rechtfertigung als ein allen anderen Versuchen überlegenes Mittel beschränken.



Es ist vollkommen unsinnig, die Qualität eines Expertensystems am Bestehen oder Verfehlen des sogenannten Turing-Tests bemessen zu wollen. Viel wichtiger ist es, ob die Trefferquote der von dem Expertensystem ermittelten Ergebnisse einem Vergleich mit den Resultaten der besten menschlichen Fachleute standhält.



Zwischen den ursprünglichen und den heutigen wissensbasierten Systemen gibt es viele Unterschiede. Einer der wichtigsten liegt in der Einbettung der Systeme in ihre Einsatzumgebung. Früher wurde fast immer von einer statischen Situation ausgegangen, deren Zustände sich während der Zeitdauer zur Bearbeitung eines Problems nur durch explizit angegebene Operationen ändern durften ("Blockwelt"). Diese unrealistische Vorstellung wird heutzutage nicht mehr vorausgesetzt. Vielmehr werden durchaus unvorhergesehene Wechselwirkungen zwischen einem wissensbasierten System und seinem Umfeld in die Betrachtungen einbezogen.

Probleme und Entwicklungen der Neuroinformatik

Prof. Dr. W. von Seelen

1. Anmerkungen zum Entwicklungsstand

Die Beschreibung eines Fachgebietes läßt sich am einfachsten durch seine Fragestellungen und Methoden gewinnen. Das zentrale Thema der Neuroinformatik ist die Entwicklung von Systemen, die lernfähig sind, eine für steigende Komplexität offene Architektur aufweisen und die in definierten Bereichen autonom zu agieren vermögen. Diese Probleme sind orientiert an biologischen Vorbildern, die die prinzipielle Lösbarkeit der genannten Fragen demonstrieren. Dieses sehr breite Interessensspektrum teilt die Neuroinformatik mit einer Reihe weiterer Fachgebiete wie Informatik, KI, oder Nachrichtentechnik. Eine spezifische Methodik wurde bisher nicht entwickelt, wohl aber ein Kanon von Verfahren, die eine enge Bindung an die Mathematik und die Physik von nichtlinearen dynamischen Systemen aufweisen. Die Entwicklung des Gebietes verlief außerordentlich rapide und - angesichts der relativ kleinen Zahl von involvierten Wissenschaftlern - insgesamt erfolgreich. Dabei zeigte sich bei der skizzierten Spannweite der Probleme, daß deren Lösung nur schrittweise zu erreichen ist. Glücklicherweise verlief die Entwicklung in der Form, daß die Teillösungen anwendungsrelevant sind und auf ein bemerkenswertes Interesse der Industrie treffen. Wie in verwandten Fachgebieten auch, ist das Interesse ungleich auf die Problembereiche verteilt und wesentlich durch die erwartete Anwendungsnähe bestimmt. Die Abbildungen 1 - 6 kennzeichnen summarisch den gegenwärtigen Stand. Sie sind einer Studie entnommen, an der die Institute für Neuroinformatik in Bochum und Bonn im Auftrag des Landes Nordrhein-Westfalen mitgearbeitet haben. Basis der Grafiken sind - aus Aktualitätsgründen - 1700 Konferenzbeiträge (1989 - 92), die Ländern und Sachgebieten zugeordnet sind. Die Abb. 2 bezieht sich auf die weltweite Aktivität, die Abb. 1 auf Europa. Die Stellung Deutschlands beruht sehr wesentlich auf den Aktivitäten, die das BMFT in zwei Programmen fördert. Japan macht gegenwärtig massive Anstrengungen, um zukünftig eine größere Rolle zu spielen. Da viele der Konferenzen in den USA stattfinden, ergab sich ein geringer natürlicher Vorteil für die Amerikaner. Die Abb. 3 und 4 kennzeichnen die thematische Verteilung und zeigen in den USA und der Bundesrepublik ein überwiegendes Interesse an Problemen des Lernens. Versucht man den

Entwicklungsstand des Gebietes in beiden Ländern zu erfassen, so fallen die meisten Arbeiten in die Testphase, in der die prinzipielle Lösbarkeit von Problemen durch das Verfahren untersucht wird. Wiederum typisch ist die stärkere Betonung theoretischer Arbeiten in Deutschland und das ausgeprägtere Bemühen um Produkte in den USA. Das Interesse am lernenden System resultiert aus der erwarteten Anwendungsnähe der Lösungen. Korrekturbedürftig ist der Tatbestand, daß in der Bundesrepublik die Zahl der Produkte unterhalb der Zeichengenauigkeit der Grafik liegen. Hierbei ist jedoch zu beachten, daß neuronale Netze bei praxisnahen Entwicklungen in Deutschland häufig wichtige Elemente eines Gesamtsystems sind.

Vergleicht man den erreichten Stand des Gebietes mit dem durch biologische Systeme demonstrierten Möglichkeiten, so liegt noch ein weites Entwicklungspotential offen. Trotzdem erscheint es erreichbar, daß in absehbarer Zeit die investierten Forschungsgelder mit den erarbeiteten Verfahren zumindest teilweise verdient werden können. Dies gilt vor allem für lernende neuronale Netze, die bei wählbarer Ausgangstopologie mit Hilfe lokaler oder globaler Lernregeln das Approximationsproblem stabil und mit hinreichender Konvergenzgeschwindigkeit lösen. Trotz großer Vielfalt der Verfahren im Detail implementieren die meisten Systeme bei Hebb-artigen Regeln ein Gradientenverfahren für die Strukturfestlegung (z.B. Backpropagation) (s. Abb.7).

Im Fall des unüberwachten Lernens nutzen die meisten Realisierungen die Kovarianzmatrix der Eingangssignale und schätzen deren Eigenvektoren. Damit sind die erreichbaren Leistungsfähigkeiten durchaus im Bereich theoretisch bekannter Systeme. Dieses Argument verdeckt jedoch die wesentliche Zielsetzung neuronaler Netze: das datengetriebene Erreichen einer guten Leistung und die Möglichkeit, sich Variationen von Randbedingungen und Zielsetzungen selbsttätig anzupassen. In vielen Fällen kann darüber hinaus ein Teil der Entwurfsarbeit eingespart werden. Mit Hilfe älterer Arbeiten aus dem Bereich der Mathematik haben neuronale Netze m.E. heute den Charakter eines Werkzeugs erreicht, das einsetzbar ist, sofern die Approximationsweite begrenzt werden kann. Trotz vieler noch möglicher Verbesserungen liegen die wichtigsten Probleme bei der Integration lernender Teilnetze in komplexere Gesamtsysteme. Diese Aufgabe sollte allerdings anwendungsbezogen gelöst werden. Entwicklungssysteme für neuronale Netze könnten diese Aufgabe erheblich erleichtern. Inwieweit Hardware-Lösungen vonnöten sind, hängt vom Anwendungsfall ab und kann angesichts der rapiden Entwicklung nicht generell entschieden werden. Eine weitere Anwendungsmöglichkeit neuronaler Netze ergibt sich ebenfalls aus ihrem guten Approximationsverhalten. Bei komplexen Systemen eignen sie sich häufig als niederdimensionale "Systembeobachter", mit deren Hilfe entweder die Daten systemgerecht zu ordnen sind (z.B. Kohonen-Netze) oder ein niederdimensionales Modell gewonnen werden kann, das steuernde

Eingriffe in das reale System begründet. In diesem Bereich sind jedoch noch Entwicklungsarbeiten zu leisten.

2. Allgemeine Probleme

Der erreichte Entwicklungsstand sollte nicht darüber hinwegtäuschen, daß die heutigen neuronalen Netze nur sehr vereinfachte "Abbildungen" von Neuronen sind und Strategien von Gehirnen allenfalls zu einem sehr kleinen Teil imitieren. Es könnte sein, daß weitergehende Lösungen davon abhängen, daß Verarbeitungs- und Kopplungsstrategien biologie-näher werden müssen. Ich möchte Sie auf einige prinzipielle Probleme hinweisen, die bedenkenswert erscheinen, da kein kleiner quantitativer Schritt, sondern ein qualitativer Sprung versucht werden muß. Verfolgt man das Ziel, lernende, autonome (intelligente) Systeme hoher Komplexität zu entwickeln und beabsichtigt man deren Einsatz in natürlichen Umgebungen, so bedürfen folgende Probleme entweder einer Lösung oder einer präziseren Formulierung.

a. Natürliche Umwelt

Eine natürliche Umwelt kann vielgestaltig sein. Sie umfaßt eine visuelle Freilandszene ebenso wie ein Gesicht oder ökologische Wirkungsgefüge. In der Regel ist sie dadurch gekennzeichnet, daß

1. die Information unscharf definiert ist,
2. Signale eine hohe aber nutzbare Redundanz aufweisen,
3. Störungen sowohl statistisch wie determiniert sein können,
4. Randbedingungen variabel sind und
5. Signalparameter sich über mehrere Zehnerpotenzen ändern können.

Da z.B. Störungen entscheidend die Struktur der Empfänger bestimmen (Punkt 3), existiert eine Lücke. Die in der Signaltheorie übliche Störunterdrückung auf der Basis von Erwartungswerten ist eine unvollständige Antwort. Auch die anderen genannten Eigenschaften des Signalraumes treffen auf kein vollständiges Instrumentarium im Bereich der Theorie.

b. Komplexität

Bei der Analyse z.B. biologischer Systeme ist deren Komplexität eine hohe Hürde. Da die Beschreibung auf Beobachtungen basiert, sind die Eigenschaften des Systems und die Komplexität der Umwelt, in der sie agieren, nicht ohne weiteres voneinander zu trennen. So basiert das Heimfindevermögen einer Ameise auf relativ einfachen Mechanismen, die Beobachtungen dagegen spiegeln im wesentlichen die Komplexität der Umwelt. Geht man davon aus, daß im Ablauf der Evolution stabile Subsysteme entstanden, die als Ganzes wiederum dem Evolutionsprozeß unterworfen wurden, so konnte nur eine spezifische Komplexität entstehen. Systeme dieser Art weisen die Eigenschaft der "Beinahe-Zerlegbarkeit" auf; d.h. die Beschreibungsmatrizen enthalten längs der Hauptdiagonalen Submatrizen, die die örtlich und zeitlich starken Wechselwirkungen beschreiben. Sie sind mit dem Restsystem durch "schwächere" Wirkungen verbunden. Die Handhabung von Komplexität ist m.E. auf derartige Systeme zu konzentrieren. Sie implizieren hierarchische Struktur und bieten möglicherweise die einzige von uns "verstehbare" Form von Komplexität. Die Physik hat die distanzabhängigen Wechselwirkungen stets zur Separierung der Probleme genutzt und z.B. bei chemischen Bindungen weder Kernkräfte noch Gravitation betrachtet. Bei informationsverarbeitenden Systemen, die konstruiert werden sollen, erscheint es sinnvoll, die Komplexität mit der Architektur zu verbinden und tragfähige Komplexitätsmaße zu definieren, - die beiden heute gebräuchlichen reichen nicht aus -.

c. Architektur

Rechner haben einen einheitlichen Aufbau (Architektur) und die wichtige Eigenschaft, daß alle Probleme, die sich auf rekursiv definierbare Funktionen abbilden lassen, zumindest prinzipiell lösbar sind. Massive praktische Restriktionen ergeben sich daraus, daß bestimmte Problemklassen (z.B. Sprachverstehen) sich nur schwer auf die gebräuchliche Rechnerarchitektur abbilden lassen. Die extreme Gegenposition geht davon aus, daß jedem Problem eine spezifische Rechnerarchitektur zuzuordnen ist. Neuronale Systeme verfügen nach heutigem Kenntnisstand über flexible Architekturen, die jeweils problemabhängig konfiguriert werden. Die Zahl der eingesetzten Grundstrukturen scheint klein zu sein. Zwei wesentliche Aspekte neuronaler Systeme sind m.E.:

1. Es existiert eine "Isomorphie" in dem Sinne, daß kleinen Veränderungen im Signalraum kleine Veränderungen in der internen Repräsentation zugeordnet werden.

2. Die Vielfalt der Leistungen basiert darauf, daß nicht in erster Linie die Prozesse, sondern die Datenformate, auf denen die Prozesse arbeiten, variiert werden. (z.B. Kohonen-Karte)

Neuronale Systeme sind - nicht notwendigerweise bestmöglich - Problembereichen angepaßt. Ihre Architektur ist ein entscheidendes Schlüsselproblem. Heutige neuronale Netze sind eine sehr unvollständige Imitation, die Leistungsbeschränkungen zur Folge haben muß.

d. Flexibilität

Durch Lernen erzeugte Flexibilität ist die wichtigste Systemeigenschaft neuronaler Netze. Zwar eröffnen sich zahlreiche Anwendungsmöglichkeiten bereits jetzt, doch weitergehende Optionen sind möglich, wenn es gelingt - es wird daran gearbeitet - die folgenden Probleme zu lösen:

1. Schnelle Implementierung vorhandenen Modellwissens
2. Nutzung von Lernstrategien in Kombination mit den heute üblichen, in der Regel lokalen, Lernregeln.
3. Nutzung der Variation von Parametern in einer reicheren Struktur, die nicht notwendigerweise Gewichte sind. Das System der Neuromodulatoren gibt zahlreiche Hinweise über Art und Ort von Parameterveränderungen.

M.E. sind weitergehende Anwendungen möglich, wenn es gelingt, die obigen Probleme in Schritten und sinnvollerweise an Anwendungsbeispielen zu lösen.

e. Redundanz

Die Nutzung von Redundanz ist nach heutigem Kenntnisstand die entscheidende Möglichkeit, die Verfügbarkeit von Systemen zu erhöhen. Informations- und Signaltheorie haben zahlreiche Methoden zur Verfügung gestellt. Bezieht man die Forderung nach Sicherheit allerdings z.B. auch auf die Kommandostruktur, die das "Verhalten" organisiert, so müssen die Probleme als weitgehend ungelöst angesehen werden.

Um komplexe Systeme einsatzfähig zu halten, muß die inhärente Systemsicherheit beliebig manipulierbar sein. Die Biologie weist hier mehrere Wege, die u.a. darauf hinauslaufen,

Signalredundanz zu transformieren und Sicherheit und Genauigkeit der Operationen austauschbar zu halten (graceful degradation).

Ein weiterer Bereich bezieht sich auf intelligente Systeme, deren Definition nach wie vor unscharf ist. Unter dem einschränkenden Aspekt der Evolutionsbiologie könnte man intelligente Systeme dadurch charakterisieren, daß sie unter variablen Randbedingungen und Gütekriterien die jeweils bestmögliche Lösung anstreben. Intelligenz wäre damit ein Ökonomieprinzip. Zur quantitativen Fassung der damit verbundenen Probleme ist allerdings noch erhebliche Forschungsarbeit zu leisten.

Die Lösung der oben nur sehr kurz skizzierten Probleme sollten weder in voller Allgemeinheit versucht werden, noch ist sie eine Voraussetzung für weitere Anwendungen. Die Evolution hat sie unter strengem Aufgabenbezug gelöst; eine vernünftige Balance zwischen Grundlagenforschung und Anwendungsentwicklungen scheint mir am vielversprechendsten zu sein. Angesichts der prinzipiellen Struktur der Probleme scheint es mir sinnvoll, die theoretischen Interessen mit denen der angrenzenden Fachgebiete zu bündeln.

3. Mögliche Forschungsthemen

Zukünftige Forschungsthemen sind der Versuch einer Prognose und damit naturgemäß irrtumsanfällig. Die nachstehend aufgelisteten Probleme fassen eine persönliche Sicht zusammen und gehen davon aus, daß die Nachbargebiete der Neuroinformatik überlappende Interessen haben. Die Themen gliedern sich in drei m.E. wesentliche und voneinander nicht unabhängige Gebiete, wobei die als "Entwicklungsbereiche" bezeichneten Probleme bereits für Spezialfälle begonnen wurden.

Wenn das Leistungspotential von Gehirnen eine Meßplatte darstellt, so erscheint mir das Architekturproblem für Systeme mit "offener" Komplexität von zentraler Bedeutung. Parallele Strukturen sind dafür eine Voraussetzung, sie sollten jedoch nicht nur unter dem Aspekt hoher Datenflüsse gesehen werden. Biologische Systeme bestechen nicht so sehr durch ihre Leistungen bei Einzelaufgaben, sondern durch ihre vielschichtige Flexibilität und Integrationsfähigkeit. Die bisher erzielten Erfolge bilden eine gute Ausgangsbasis. Der mit Autonomie (Intelligenz) bezeichnete Bereich muß sorgsam strukturiert werden. Laute Versprechungen entbehren angesichts der fundamentalen Probleme einer vernünftigen Grundlage. Andererseits sind die Optionen derartiger Systeme, die die Evolution als Verwalterin von Mängeln hervorgebracht hat, außerordentlich weitreichend.

Forschungsbereiche

Architektur f(A,S,P)	Flexibilität	Autonomie (Intelligenz)
1. Senso-Motorische Systeme mit "offener Komplexität"	1. Entwicklung von Lernstrategien	1. Effiziente "Match"-Prozesse auf unterschiedlichen Niveaus
2. Multi-Sensor-Systeme für "natürliche" Umwelten (Repräs.)	2. Strukturabhängige günstige Lernverfahren (ev. Parameter)	2. Hierarchisierung von Problemen
3. Verhaltensorganisierende Systeme mit "Kommando"-Redundanz	3. Problemabhängige Kodierung von Wissen	3. Schnelle mehrdimensionale Optimierung variabler Güter
4. Systeme wählbarer Sicherheit (Redundanz)	4. Schnelle Modelladaption	4. Kombinationen gestufter Optimierung variabler Güter
5. Problemabhängige Architekturen	5. Verfahren für präzise Approximation	
6. Kodierung von Wissen in hierarch. Systemen		
7. "Evolutionäre" Strukturoptimierung		

Biologie-nahe Forschung

1. Eigenschaften geschichteter Strukturen (funkt. Repräsentation)
2. Interpretierende "Kontexte" in frühen Verarbeitungsstufen (MODELLE)
3. Analyse senso-motorischer Verhaltensabläufe (Kompensationsverfahren)
4. Aufarbeitung phylogenetischer, autogenetischer und neuroethologischer Aspekte (Konstruktionsregeln)

Entwicklungsbereiche

1. Analyse nicht zu komplexer Signale aus natürl. Umwelt für Überwachung, Fehlerbestimmung usw.
2. Adaptive Regelungssysteme für Prozeßsteuerung und Roboter
3. Lernende Mustererkennung bei begrenzten Klassenzahlen
4. Aktive sensorische Systeme, die den Datenfluß abhängig von der Informationsverarbeitung steuern
5. Kopplung KI-basierter Systeme mit neuronalen Netzen
6. Erzeugung angepaßter Datenformate (z.B. Kohonen-Karten) für effizientes Lernen
7. "Aufmerksamkeits"-gesteuerte Systeme

Anhang: Abbildungen

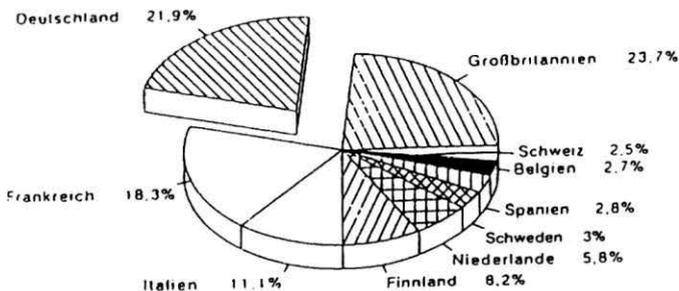


Abb.1: Aktivitäten in Form von Konferenzpublikationen für Europa

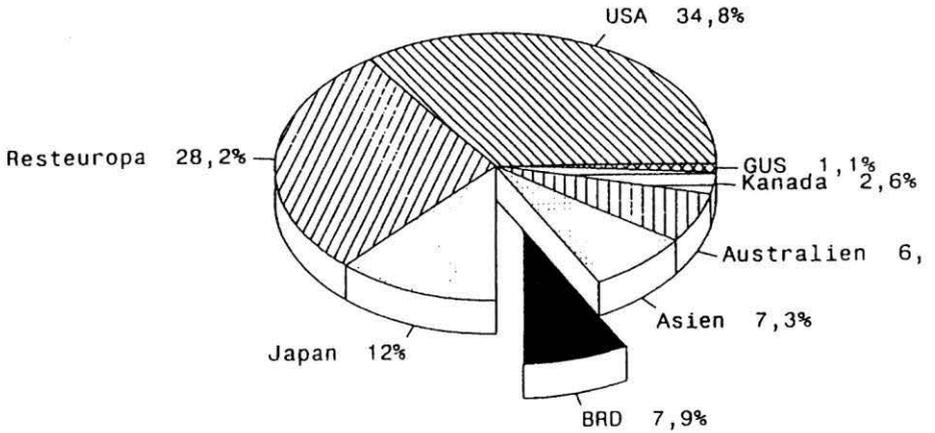


Abb.2: Aktivitäten in Form von Konferenzpublikationen weltweit

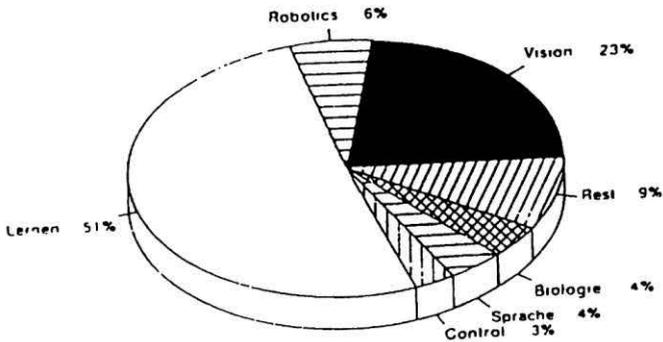


Abb.3: Themenverteilung in den USA

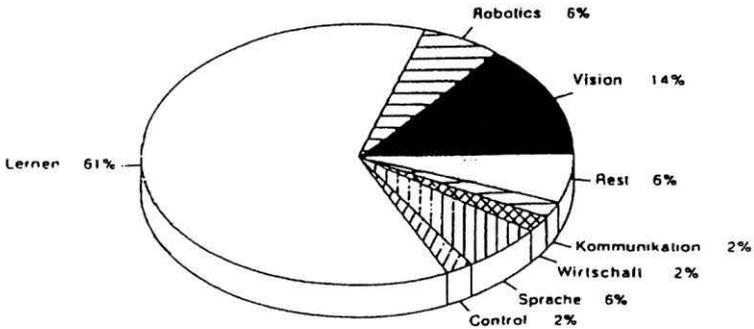


Abb.4: Themenverteilung in der Bundesrepublik

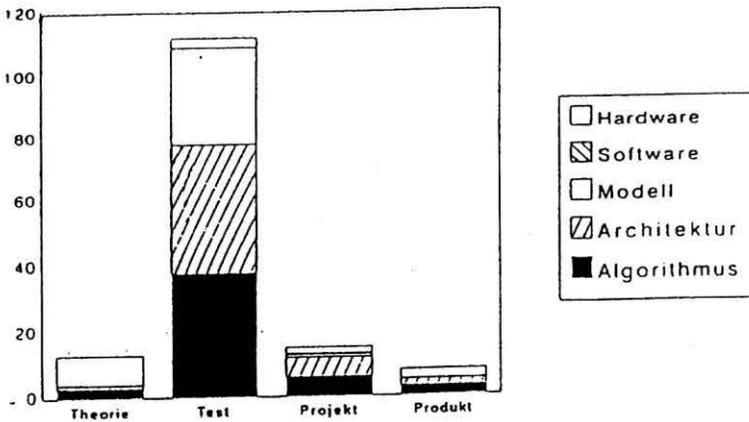


Abb.5: Entwicklungsstand der Arbeiten in den USA

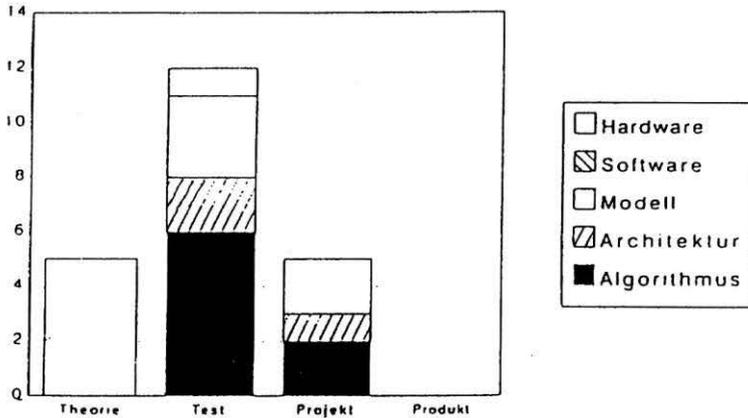


Abb.6: Entwicklungsstand in der Bundesrepublik

	unüberwacht	↔	überwacht
lokal	Hebb-Regel: $\Delta w_{ij} = \lambda e'_i e_j$ dto. mit „Vergessen“ $\Delta w_{ij} = \lambda e'_i (e_j - w_{ij})$ Kovarianz-Lernen: $\Delta w_{ij} = \lambda (e'_i - \bar{e}') (e_j - \bar{e})$		Verstärkungslernen: $\Delta w_{ij} > 0$ 1. e'_i war erregt 2. e_j hat beigetragen 3. Ergebnis wird besser Widrow-Hoff: $\Delta w_{ij} = \lambda (t_i - e'_i) e_j$
↕	Wettbewerbslernen: $\Delta w_{ij} = \lambda e'_i e_j / (\sum w_{ij}^2)$		Back-Propagation
global	Vorbereitung der Gewichte aufgrund analytischer Betrachtungen (z.B. Pseudo-Inverse)		

Abb.7: Schema von Lernverfahren

ω charakterisiert die Gewichte e_j , die präsynaptische e' , die postsynaptische Erregung und t die gewünschte Ausgangsgröße.

Das Backpropagationsverfahren repräsentiert ein Gradientenverfahren und ist für wichtige Fälle der Nutzung der Widrow-Hoff-Regel äquivalent.

Möglichkeiten und Grenzen der Künstlichen Intelligenz - Überlegungen aus der Sicht der Automatisierungstechnik

Prof. Dr.-Ing. J. Lunze

1. Gegenstand der Automatisierungstechnik

1.1 Automatisierungsaufgaben

Gegenstand der Automatisierungstechnik sind dynamische Prozesse, deren Verhalten im Sinne einer gegebenen Zielstellung beeinflusst werden soll. Bei den Prozessen kann es sich um verfahrenstechnische Anlagen, Kraftwerke oder Fertigungszellen handeln. Im folgenden soll untersucht werden, welche neuen Möglichkeiten für die Automatisierungstechnik durch den Einsatz von Methoden der Künstlichen Intelligenz (KI) eröffnet werden.

In der Sprache der Regelungs- und Automatisierungstechniker wird der dynamische Prozeß in einem Blockschaltbild dargestellt, in dem Pfeile Signale und Blöcke Signalverarbeitungseinheiten kennzeichnen. In Abbildung 1 sind der Prozeß sowie die Automatisierungseinrichtung durch je einen derartigen Block repräsentiert. Die Meßgrößen kennzeichnet ein Pfeil vom System zur Automatisierungseinrichtung, die Stellgrößen ein Pfeil in umgekehrter Richtung. Umwelteinflüsse, die im Sinne der Automatisierungstechnik Störungen darstellen, sind durch einen auf den Prozeß gerichteten Pfeil dargestellt.

Die Automatisierung beschäftigt sich mit der Aufgabe, zu einem gegebenen dynamischen System eine Automatisierungseinrichtung zu schaffen, so daß das dynamische System und die Automatisierungseinrichtung gemeinsam durch den menschlichen Operator gesteuert werden können. Dabei sind mehrere Typen von Aufgaben zu unterscheiden (Tabelle 1).

Die Tabelle gibt einen Einblick in die Aufgaben, die durch die Automatisierung gelöst werden müssen. Offensichtlich beziehen sich die oberen Zeilen auf Probleme, die schnell und mit genauen Meßinformationen über den Prozeß gelöst werden müssen, während nach unten der betrachtete Zeithorizont und die Größe des betrachteten Teiles der gesamten Anlage zunehmen und die Genauigkeit der Meßinformationen abnehmen. Die rechte Spalte gibt an, ob diese Aufgaben typischerweise vor allem automatisch durch die Automatisierungseinrichtung übernommen werden oder ob sie manuell durch einen menschlichen Operator ausgeführt

werden müssen. Häufig sind Mensch und Maschine zu unterschiedlichen Teilen an der Lösung der Aufgabe beteiligt.

Gegenstand der Automatisierungstechnik

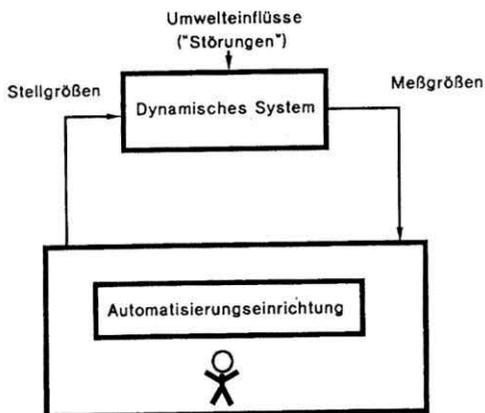


Abbildung 1: Grundstruktur einer Automatisierungslösung

Automatisierungsaufgabe	Funktion	Beispiele	Ausführung
Meßwertverarbeitung	Vorverarbeitung der durch Meßrauschen gestörten Meßwerte	Ermittlung des Frequenzspektrums; Trendanalyse	automatisch
Steuerung und Regelung	kontinuierliche oder quasi-kontinuierliche Vorgabe der Stellgröße, so daß die Meßgrößen vorgegebene zeitliche Verläufe erhalten	Fahrstuhlsteuerung; Regelung eines Behälterstandes	automatisch
Prozeßüberwachung	Kontinuierliche Kontrolle wichtiger Prozeßsignale auf Einhaltung vorgegebener Schranken	Grenzwertüberwachung für Temperatur und Druck	automatisch oder manuell
Prozeßführung	Steuerung komplexer Teilprozesse entsprechend vorgegebener Sollwerte oder Wertefolgen für wichtige Signale	An- und Abfahren von verfahrenstechnischen Reaktoren	manuell
Betriebsführung	Steuerung des Produktionsbetriebes durch Vorgabe von Sollwerten für einzelne Teilprozesse	Bestimmung optimaler Spannungswerte für ein Elektroenergienetz	manuell
Betriebsplanung	Festlegung von langfristigen Zielen für einen Produktionsbetrieb	Produktionsplanung: Festlegung von Soll-Stückzahlen	manuell

1.2 Charakteristik gegenwärtiger Lösungsmethoden

Im folgenden wird eine allgemeine Charakterisierung der Methoden angegeben, die heute vorrangig für die Lösung der in Tabelle 1 aufgeführten Automatisierungsaufgaben eingesetzt werden. Diese Charakterisierung soll helfen, die durch KI-Methoden behandelbaren Probleme einzugrenzen. Dabei ist zunächst herauszustellen, daß sich alle Automatisierungsaufgaben durch folgende Elemente beschreiben lassen:

- **Modelle** des dynamischen Prozesses: Diese Modelle repräsentieren das *Wissen über die dynamischen Eigenschaften* des zu automatisierenden Prozesses. Für einen gegebenen aktuellen Prozeßzustand und die möglicherweise eingesetzten Steuerungen kann mit diesem Modell das zukünftige Verhalten des Systems vorhergesagt und bezüglich der Zielstellung bewertet werden.
- Der **aktuelle Prozeßzustand**: Die Meßgrößen kennzeichnen die Lage des Systems im Zustandsraum, also z.B. die aktuellen Temperatur- und Druckwerte. Sie stellen den *Kontext* dar, in dem die Automatisierungsaufgabe zum gegenwärtigen Zeitpunkt gelöst werden muß.
- Die **Zielstellung**: Durch sie wird beschrieben, welchen Bedingungen das Verhalten des dynamischen Systems genügen soll.

Um diese Aufgaben automatisch, z.B. mit Hilfe eines Rechners, lösen zu können, muß das betrachtete Automatisierungsproblem in ein exakt formuliertes (mathematisches) Problem überführt werden. Die verwendeten Lösungsalgorithmen hängen in entscheidendem Maße von der Art der verwendeten Modelle und der Art der Formulierung des Automatisierungszieles ab.

In dieser Hinsicht ist die Automatisierungstechnik gegenwärtig durch folgende Merkmale gekennzeichnet:

- Es werden quantitative Modelle wie z.B. Differentialgleichungen, Automaten oder Petri-netze verwendet, um das Verhalten des dynamischen Prozesses mit bzw. ohne Automatisierungseinrichtung zu beschreiben.
- Es wird mit quantitativen Zustandsinformationen gearbeitet, d.h., die Meßgrößen sollen den Prozeßzustand so genau wie möglich wiedergeben.

- Das Automatisierungsziel soll exakt festgelegt werden, z.B. durch eine Sollkurve, auf dem ein Roboterarm entlangfährt, oder feste Werte, auf denen Drücke und Temperaturen einer verfahrenstechnischen Anlage unabhängig von äußeren Störungen gehalten werden sollen.
- Die Automatisierungseinrichtung verarbeitet die Meßwerte numerisch und berechnet daraus die erforderlichen Stelleingriffe.

1.3 Ziele zukünftiger Automatisierungslösungen

Vorrangige Ziele künftiger Automatisierungslösungen bestehen in der Erhöhung der Autonomie technischer oder nichttechnischer Anlagen und in der Verbesserung der Mensch-Maschine-Interaktion.

1.3.1 Erhöhung der Autonomie

Unter der Autonomie einer technischen Anlage versteht man die Fähigkeit der Anlage, das Verhalten den aktuellen Umweltbedingungen anzupassen. Auf Automatisierungsaufgaben bezogen sind dafür zwei Probleme zu lösen:

Erstens muß die Automatisierungseinrichtung erkennen, welche Probleme zum aktuellen Zeitpunkt vorrangig gelöst werden müssen. Der Pfeil für die Störungen in Abbildung 1 zeigt, daß das Verhalten des dynamischen Systems durch die Umweltbedingungen beeinflußt wird. Deshalb ist auch von der Umwelt abhängig, welche Automatisierungsaufgabe vorrangig zu lösen ist und wie diese Aufgabe konkret aussieht. So muß z.B. bei autonomen Fahrzeugen aus einer Videoaufnahme der Umwelt abgeleitet werden, wie die Fahrtroute zu wählen ist bzw. wo Hindernisse zu umfahren sind. Die dabei ermittelte Fahrtroute schreibt dann die Sollgrößen für die Fahrzeugregelung vor. Sie ist nicht von vornherein vorgegeben, sondern muß von dem Fahrzeug selbst ausgewählt werden. Das Fahrzeug arbeitet in dieser Hinsicht autonom.

Zweitens müssen die vorhandenen Modelle, Zustandsinformationen und Automatisierungsziele bezüglich ihrer Wichtigkeit für die aktuell zu lösenden Automatisierungsaufgaben bewertet werden. So muß für die Auswahl der Fahrtroute ein Lageplan bzw. eine aktuelle Übersicht über Hindernisse verwendet werden, aber es ist für diese Aufgabe gleichgültig, ob die Motorentemperatur im vorgegebenen Sollbereich liegt oder nicht. Die situationsabhängige Auswahl der für die Lösung einer Aufgabe relevanten Informationen ist ein wesentliches Element bei der Realisierung autonomer Systeme.

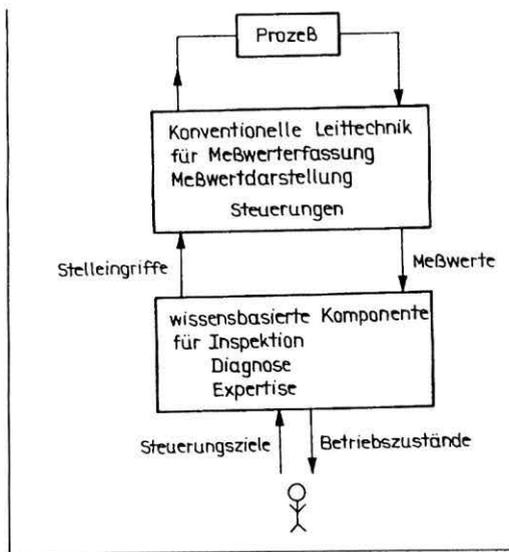


Abbildung 2: Intelligente Leittechnik

1.3.2 Mensch-Maschine-Interaktion

Typisch für das Niveau der Mensch-Maschine-Kommunikation heutiger Automatisierungslösungen ist, daß der menschliche Operator eine Vielzahl von einzelnen Meßwerten angezeigt bekommt. Die dafür notwendige Informationserfassung, -aufbereitung und -darstellung wird durch die Leittechnik übernommen (Abbildung 2). Zur Verbesserung der Übersicht sind diese Informationen in Fließbildern angeordnet, so daß auch graphisch ein direkter Bezug zwischen dem Meßwert und dem Ort, an dem die betreffende Meßgröße aufgenommen wurde, hergestellt wird.

Zukünftig soll die Kommunikation jedoch auf semantischem Niveau ablaufen. Der menschliche Operator möchte der Automatisierungseinrichtung Problembeschreibungen und Lösungsmethoden mitteilen anstelle die zur Erzielung eines vorgegebenen Arbeitspunktes notwendige Folge von Stelleingriffen Schritt für Schritt vorzugeben. Abbildung 2 zeigt, daß sich das Abstraktionsniveau der Kommunikation zwischen Operator und Leittechnik verändert: Anstelle von einzelnen Stelleingriffen werden Steuerungsziele vorgegeben; anstelle einer Vielzahl von Meßwerten erhält der Operator eine globale Darstellung und Bewertung des aktuellen Betriebszustandes.

2. Erweiterung der Automatisierung durch Künstliche Intelligenz

2.1 Problemklassen, die mit Hilfe der Künstlichen Intelligenz gelöst werden können

Der Einsatz von Methoden der Künstlichen Intelligenz in der Automatisierungstechnik ist ein Gebiet, auf dem seit mehreren Jahren intensiv geforscht wird. Das BMFT, die Deutsche Forschungsgemeinschaft und mehrere Europäische Programme unterstützen dies. Dabei geht es vor allem um Aufgabenklassen, die mit konventionellen Methoden der Automatisierungstechnik nicht gelöst werden können, bei denen es durch Verwendung der Künstlichen Intelligenz aber gelingen kann, die Problemstellung rechnergestützt zu lösen. Dabei kristallisieren sich drei große Problemklassen heraus, die nachfolgend behandelt werden:

2.1.1 Einführung neuer Ausdrucksformen und Verarbeitungsprinzipien

Durch Methoden der Wissensrepräsentation und Wissensverarbeitung können neue Formalismen für die Darstellung und Verarbeitung von Wissen über dynamische Systeme und über die Lösung von Automatisierungsaufgaben eingeführt werden. Die folgenden Beispiele verdeutlichen diese Entwicklungsrichtungen:

- Die Wissensdarstellung in Form von Regeln, logischen Formeln, Frames oder Constraints - um nur einige zu nennen - ermöglichen es, diskrete Objekte, Zustände und Erscheinungen sowie deren Verknüpfungen in einer verarbeitbaren Form zu erfassen. Diese Formen können in der Automatisierungstechnik vor allem dort eingesetzt werden, wo Differentialgleichungen oder Automaten als Repräsentationsformen nicht anwendbar sind.
- Die Verarbeitung *vagen Wissens* wird durch die Verwendung entsprechender Repräsentationsformen erleichtert. Ein aktuelles Beispiel ist das Gebiet der Fuzzy-Regelung, bei dem das Reglergesetz in Form von WENN-DANN-Regeln formuliert wird (Abbildung 3, [2]). Wenn der Operator ausreichend Erfahrungen mit der zu regelnden Anlage besitzt und deshalb in Form derartiger Regeln vorgeben kann, wie der Regler arbeiten soll, so erleichtern Fuzzy-Regler die Implementierung dieser Verhaltensregeln.

Um dabei der Denkweise des Operators entgegenzukommen, werden die Meßgrößen bezüglich ihrer Größe qualitativ bewertet ("fuzzifiziert"). Das heißt, anstelle des genauen Meßwertes wird nur mit der Bewertung des Meßwertes durch linguistische Werte wie z.B. "groß", "mittelgroß" oder "klein" gearbeitet. Da der Regler einen quantitativ exakt bestimmten Wert für die Stellgröße vorgeben muß, steht am Ende der Verarbeitung der Schritt der

"Defuzzifizierung", durch den von den linguistischen Werten wieder zu Zahlenwerten übergegangen wird.

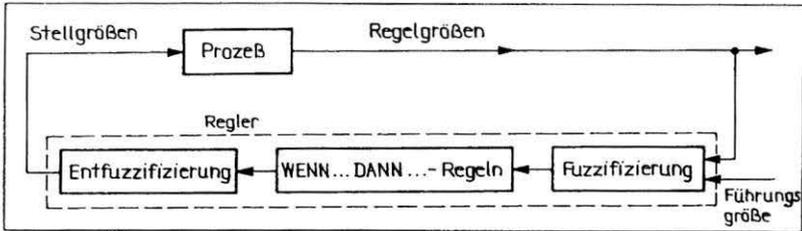


Abbildung 3: Fuzzy-Regelung

- Die Fähigkeit des *Lernens* kann mit den in der Künstlichen Intelligenz erarbeiteten Prinzipien viel umfangreicher in die Automatisierungstechnik eingeführt werden, als dies durch adaptive oder lernende Regelungen bisher möglich war [8]. Die Automatisierungseinrichtung verbessert die Steuerung, da sie "aus Erfahrungen lernt". Frühere Steuereingriffe werden bezüglich der mit ihnen erzielten Güterwerte der Regelung oder Steuerung beurteilt und gegebenenfalls im Falle eines erneuten Auftretens der betreffenden Situation verbessert.

Besonderes Interesse ruft deshalb derzeit die Anwendung von künstlichen neuronalen Netzen in der Automatisierungstechnik hervor. Für die Klassifikation von Betriebszuständen ist mit diesen Netzen bereits erfolgreich gearbeitet worden; für die Anwendung zur Steuerung dynamischer Systeme müssen die Netze jedoch noch für zeitabhängige Daten erweitert werden.

Diese Beispiele zeigen, daß die Künstliche Intelligenz zur Erweiterung des Spektrums derjenigen Automatisierungsaufgaben beitragen kann, die durch den Rechner automatisch gelöst werden können. Dies drückt sich in einer Verschiebung der Grenzlinie aus, die in Tabelle 1 die automatisch gelösten von den manuell gelösten Automatisierungsaufgaben trennt.

2.1.2 Kognitiv adäquate Repräsentationen für die Mensch-Maschine-Kommunikation

Die Verwendung von Methoden der Wissensrepräsentation ermöglicht die Verbesserung der Mensch-Maschine-Kommunikation vor allem in bezug auf das Abstraktionsniveau, auf dem sich dieser Dialog vollzieht. Dies wird an folgenden Beispielen deutlich:

- Die Vorgaben des Automatisierungszieles soll nach Möglichkeit in einer dem Fachmann verständlichen Sprache und auf einem Abstraktionsniveau erfolgen, auf dem der Fachmann gewöhnlich seine Probleme löst. Am Beispiel der Robotersteuerung bedeutet dies, daß der Fachmann die Montageaufgabe dem Roboter als Aufgabenstellung oder als Folge von zu lösenden Teilschritten mitteilen will und nicht - wie bisher - durch detaillierte Vorgabe der vom Roboterarm auszuführenden Bewegungen. In [7] ist eine Robotersteuerung beschrieben, bei der in einer Wissensbasis die Beschreibungen der Teilschritte einer Montage sowie ein Umweltmodell des Roboters abgelegt sind. Für eine gegebene Montageaufgabe kann der Roboter aus diesem Wissen die Folge der Montageschritte selbständig ableiten und die für den Montageprozeß notwendigen Steuerungen des Armes und des Greifers im Detail berechnen. Die Umsetzung der Montageaufgabe in die einzelnen Steuerungsaufgaben erfolgt also automatisch.
- Die Künstliche Intelligenz ermöglicht es, die Kommunikation zwischen Mensch und Maschine mit Hilfe neuer Kommunikationsmittel herzustellen. Die Sprachverarbeitung, dreidimensionale Graphiken oder die Szenenanalyse sind Beispiele, die verdeutlichen, daß die Automatisierungseinrichtung selbständig die aktuelle Betriebssituation aus Bildern oder gesprochenen Anweisungen erkennen und Steuervorschläge oder Situationsbeschreibungen in Form von Bildern oder Sprache ausgeben kann.

2.1.3 Neue Rechnerarchitekturen und neue Softwaretechnologie

Die Implementierung komplexer Automatisierungsverfahren wird durch neue Rechnerarchitekturen und durch die Softwaretechnologie der Künstlichen Intelligenz unterstützt. Parallel arbeitende Rechner sind nicht nur zur Gewährleistung der Echtzeitverarbeitung notwendig, sondern für die Betrachtung dynamischer Prozesse mit i.allg. vielen Teilprozessen auch der Aufgabenstellung adäquat. In der Parallelität des Rechnernetzes spiegelt sich die Parallelität wieder, mit der die Teilprozesse der technischen Anlage ablaufen und in ihrer Gesamtheit das Verhalten des Prozesses bestimmen.

Die neue Softwaretechnologie hat neben der Implementierung von einzelnen Verfahren auch in der Gestaltung der Automatisierungseinrichtung ihren Niederschlag gefunden. Moderne Leitsysteme werden objektorientiert aufgebaut, d.h. anstelle langer, unübersichtlicher Listen von Signalen und Grenzwerten stehen Module für die kompakte Beschreibung einzelner Meßgrößen. Die hierarchische Anordnung der Meßgrößen in Klassen von Temperatur-, Druck- oder Durchflußgrößen erleichtert die Definition der einzelnen Attribute dieser Meßwerte und die Verwendung dieser Meßwerte zur Beschreibung des aktuellen Betriebszustandes. Auf diese Weise werden moderne Leitsysteme völlig neu gegliedert.

2.1.4 Charakterisierung zukünftiger Automatisierungslösungen

Im Vergleich zu der im Abschnitt 1.2 angegebenen Bewertung der gegenwärtigen Automatisierung können zukünftige Lösungswege folgendermaßen charakterisiert werden:

- Es werden qualitative Modelle verwendet, um das Verhalten des dynamischen Prozesses mit bzw. ohne Automatisierungseinrichtung darzustellen. Diese Modelle beziehen sich typischerweise auf symbolische (anstelle von numerischen) Größen, stellen den dynamischen Prozeß auf einem höheren Abstraktionsniveau dar und nehmen keine Rücksicht auf Details der zeitlichen Entwicklung der einzelnen Meßgrößen.
- Es wird mit qualitativen Zustandsinformationen gearbeitet, d.h., die Meßgrößen werden bezüglich ihrer Größe klassifiziert, aber nicht quantitativ präzise weiterverarbeitet. Ein typisches Beispiel sind Alarmmeldungen, die lediglich anzeigen, daß ein Signal eine vorgegebene Schranke überschritten hat, aber nicht zu erkennen geben, welchen exakten Wert die betreffende Größe hat.
- Die Automatisierungsaufgabe wird durch mehrere, sich teilweise widersprechende Zielstellungen beschrieben. Diese Ziele haben im Vergleich untereinander zeitlich wechselnde Prioritäten, so daß die Automatisierungsaufgabe durch Veränderung der Umwelteinflüsse bzw. des Prozeßzustandes ständig modifiziert wird.
- Die Automatisierungseinrichtung verarbeitet die qualitativen Modelle und Zustandsinformationen symbolisch. Es dominieren Entscheidungsprobleme gegenüber den Berechnungsaufgaben der bisherigen Automatisierungslösungen.

2.2 Einsatzbedingungen der Künstlichen Intelligenz in der Automatisierungstechnik

Die Frage, wieweit die Künstliche Intelligenz die Automatisierungstechnik beeinflussen kann, ist entscheidend durch die Einsatzbedingungen bestimmt, unter denen KI-Methoden heute und in Zukunft in der Automatisierungstechnik zur Anwendung gebracht werden können. Die im folgenden aufgeführten Bedingungen sind typisch für die Anwendung von KI-Methoden auf dynamische Systeme:

2.2.1 Konfrontation des Menschenbildes der Ingenieure mit dem Menschenbild der Künstlichen Intelligenz

Die Rollenverteilungen, die der Automatisierungstechniker und der Informatiker zwischen Rechner und Mensch vornehmen, sind sehr unterschiedlich. Der Ingenieur orientiert sich an der technischen Lösbarkeit einer gegebenen Aufgabe, unabhängig davon, wie der Lösungsweg aussieht. Für ihn steht der Erfolg, d.h. die Lösung der Aufgabe, im Mittelpunkt. Da eine Vielzahl von Automatisierungsaufgaben bereits ohne die Künstliche Intelligenz gelöst wurden, ist eine Skepsis gegenüber dem häufig geäußerten, sehr hohen Anspruch der Künstlichen Intelligenz nicht zu übersehen. Die Künstliche Intelligenz wird deshalb erst dann zu einem breiteren Einsatz in der Automatisierungstechnik kommen, wenn an prototypischen Anwendungen nachgewiesen wurde, daß praxisrelevante Problemstellungen, die der bisherigen Automatisierung verschlossen blieben, mit Hilfe der Künstlichen Intelligenz zuverlässig gelöst werden können.

2.2.2 Integration von systemtheoretischen Verfahren und KI-Methoden

Es ist unabdingbar, die KI-Methoden in die vorhandenen Automatisierungslösungen zu integrieren. In der Automatisierungstechnik sind viele Probleme bereits gelöst - im Gegensatz zu anderen Anwendungsgebieten der KI. Neue Lösungen müssen auf diesen "alten" Lösungen aufbauen. Dies wird vor allem deutlich, wenn man beachtet, daß die Prozeßkopplung der Automatisierungseinrichtung in der Mehrzahl der Fälle so bleiben wird, wie sie bereits ist. Das heißt, es ist notwendig, die systemtheoretischen Herangehensweisen der Regelungstechnik mit den qualitativen Betrachtungen der Künstlichen Intelligenz zu verknüpfen. Dies ist nicht nur eine Aufgabe der Kopplung von numerischer und symbolischer Informationsverarbeitung. Grundlegende Begriffe wie die des Zustandes eines dynamischen Systems müssen mit ähnlichen Begriffen der Künstlichen Intelligenz in Beziehung gesetzt werden [3], [6].

2.2.3 Wissensverarbeitung unter Echtzeitforderungen

Die Automatisierungseinrichtung arbeitet - besonders auf der in Tabelle 1 in den oberen Zeilen eingetragenen Teilproblemen - unter Echtzeitbedingungen. Da viele Methoden der Künstlichen Intelligenz Suchprobleme zum Inhalt haben, stellt die Echtzeitfähigkeit dieser Methoden eine erhebliche Schranke für die gegenwärtige Anwendungsbreite dar. Es ist jedoch möglich, unter Nutzung spezifischer Eigenschaften dynamischer Systeme wie z.B. der kausalen Struktur, Wissensverarbeitungsmethoden zu entwickeln, die für das betrachtete Anwendungsgebiet besonders effektiv sind. Ein Beispiel für eine derartige Weiterentwicklung von KI-Methoden für automatisierungstechnische Aufgaben ist in [4] beschrieben. Dieses und andere Beispiele machen deutlich, daß die Automatisierungstechnik nicht nur ein Anwendungsgebiet für die Künstliche Intelligenz darstellt, sondern spezifische Forschungsarbeiten notwendig sind, um die auf dem Gebiet der Künstlichen Intelligenz für einen sehr breiten Einsatz entwickelten Methoden für die spezifischen Gegebenheiten des Anwendungsgebietes "Automatisierungstechnik" aufzubereiten.

3. Grenzen der Künstlichen Intelligenz in der Automatisierungstechnik

Die fortschreitende Forschung wird die Grenzen für den Einsatz von KI-Methoden auf dem Gebiet der Automatisierungstechnik ständig weiter hinausschieben. Aus heutiger Sicht sind es aber vor allem drei Gründe, die eine zu optimistische oder gar euphorische Überbewertung dessen, was Künstliche Intelligenz in Zukunft möglich machen wird, dämpfen:

- **Kreativitätsschranke:** Die Künstliche Intelligenz ermöglicht es, Teile menschlicher Denk- und Entscheidungsprozesse zu formalisieren und dem Rechner zu übertragen. Man ist sich allerdings einig: Kreativität kann nicht durch Künstliche Intelligenz ersetzt oder nachgebildet werden. In der Automatisierungstechnik heißt das, daß die Ziele der Automatisierung vorgegeben werden müssen. Auch wenn es - wie im Beispiel der Robotersteuerung - gelingt, von sehr abstrakt formulierten technologischen Zielstellungen die Regelungs- und Steuerungsalgorithmen automatisch abzuleiten, so muß dennoch die Gesamtaufgabe der Automatisierung zunächst einmal vorgegeben und strukturiert werden. Der Mensch bleibt Bestandteil des Entwurfsprozesses und des Betriebes von Automatisierungsanlagen.
- **Wissensakquisitionsschranke:** Die Lösung einer Automatisierungsaufgabe ist an die Verfügbarkeit von Modellen und Zustandsinformationen gebunden. Diese Modelle können auch durch Wissenserwerb aus den Erfahrungen gewonnen werden, die der menschliche

Operator mit der Anlage gemacht hat. In der Regelungstechnik ist der Weg typisch, bei der eine eingehende theoretische oder experimentelle Analyse der zu steuernden Anlage zum Prozeßmodell führt. In jedem Fall muß Wissen über den Prozeß in Form von Modellen (die auch die Form von Wissensbasen haben können) vorliegen, bevor eine Automatisierungsaufgabe gelöst werden kann. Die zweite wichtige Voraussetzung ist die Verfügbarkeit aktueller Zustandsinformationen des Prozesses. Je komplizierter die automatisch zu lösende Aufgabe ist, umso mehr aktuelle Informationen sind für ihre Lösung erforderlich.

- **Schranke der Verantwortbarkeit:** Die Künstliche Intelligenz soll eingesetzt werden, um so viele Lösungsschritte einer Automatisierungsaufgabe wie möglich dem Rechner zu übertragen. Dieses Bestreben wird durch die Frage begrenzt, welche Rollenverteilung ingenieurtechnisch verantwortet werden kann. Dabei geht es nicht nur um eine juristische, sondern auch um eine ökonomische und ökologische Verantwortung.

Ein wichtiges Problem besteht im Testen von KI-Systemen. Viele KI-Lösungen können heute für Anlagen mit erhöhtem Sicherheitsrisiko nicht eingesetzt werden, weil es keine überzeugenden Verfahren zum Nachweis der Fehlerfreiheit von KI-Programmen gibt. Dabei spielt eine entscheidende Rolle, daß die meisten KI-Programme mit der Suche einen Prozeß mit unvorhersehbarem und nicht im Detail untersuchten Lösungsweg enthalten. Ein weiteres Problem ist die Zuverlässigkeit der Mensch-Maschine-Interaktion. Beim Übergang von automatisch zu manuell gelösten Automatisierungsaufgaben ist es notwendig, den Operator so zu informieren, daß er die Konsequenzen seiner Steuerhandlungen erkennen kann und die Anlage nur in den Situationen sich selbst überläßt, die die Automatisierungseinrichtung mit Sicherheit allein bewältigen kann.

4. Thesen über die Möglichkeiten und Grenzen der Künstlichen Intelligenz im Ingenieurbereich

Abschließend soll der Versuch unternommen werden, die für die Automatisierungstechnik exemplarisch abgeleiteten Bewertungen über den Einsatz der Künstlichen Intelligenz für das Gebiet der Ingenieurwissenschaften zu verallgemeinern. Die folgenden Thesen sind nicht als Abschluß, sondern als Beginn der Diskussion über die Verantwortbarkeit der Künstlichen Intelligenz gedacht.

1. Künstliche Intelligenz erweitert die Ausdrucksmöglichkeiten und Verarbeitungsprinzipien für ingenieurtechnisches Fachwissen und führt zu einer wesentlichen Erweiterung derjeni-

gen Aufgabenklassen, die rechnergestützt lösbar sind. Sie ermöglicht eine "Intelligenzerhöhung der Technik".

2. Künstliche Intelligenz verbessert die Kommunikationsmöglichkeiten zwischen Mensch und technischer Anlage.
3. Künstliche Intelligenz unterstützt den Entwurf, die Konstruktion und die Inbetriebnahme technischer Anlagen und beschleunigt dadurch den Innovationszyklus.
4. Die Anwendung der Künstlichen Intelligenz ist beschränkt auf die Lösung vorgegebener ingenieurtechnischer Aufgabenstellungen ("Kreativitätsschranke").
5. Der Einsatz der Künstlichen Intelligenz ist an die Schaffung der informationstechnischen Infrastruktur von KI-Systemen gebunden. Dies betrifft sowohl die Versorgung von KI-Systemen mit Modellen über den betrachteten Prozeß bzw. Sachverhalt als auch die Bereitstellung von aktuellen Meßgrößen, die die gegenwärtige Situation bzw. den gegenwärtigen Prozeßzustand beschreiben. ("Wissensakquisitionsschranke")
6. Die Frage nach der Verantwortbarkeit des Einsatzes von KI-Methoden im ingenieurtechnischen Bereich ist Bestandteil der viel allgemeineren Frage, inwieweit der Einsatz technischer Verfahren bzw. Anlagen aus ökonomischer, ökologischer, juristischer, moralischer oder kultureller Sicht verantwortet werden kann.

Literatur

Föllinger, O.: Regelungstechnik, Hüthig-Verlag, Heidelberg, 1992.

Kiendl, H. u.a.: Fuzzy Control, Automatisierungstechnik **41** (1993) H.1 ff.

Lunze, J.: The notion of the state in systems theory and artificial intelligence, Intern. J. Intelligent Systems Engineering, 1994.

Lunze, J.; Schiller, F.: Logikbasierte Prozeßdiagnose unter Nutzung der kausalen Struktur dynamischer Systeme, Automatisierungstechnik **40** (1990) H.2, 54-58, H.3, 104-109.

Lunze, J.; Schwarz, W.: Künstliche Intelligenz, Verlag Technik, Berlin, 1990.

Reinisch, K.: Kybernetik und Informatik - gemeinsame Entwicklungsstufen, wissenschaft und fortschritt **38** (1988) H.6, 156--159.

Simon, W.: Untersuchungen zu einer intelligenten und lernfähigen Robotersteuerung für die Montageautomatisierung, VDI-Fortschritt-Berichte, Reihe 20, Nr. 47, VDI-Verlag, Düsseldorf 1991.

Tolle, H.; Ersü, E.: Neurocontrol, Springer-Verlag, Berlin, 1992.

Leitbilder der Technik - ihre Bedeutungen, Funktionen und Potentiale für den KI-Diskurs

Prof. Dr. M. Dierkes

1. KI-Diskurs: Das Projekt - Faszination, Machbarkeit und Wünschbarkeit

1.1 Die Faszination oder "there's no business like ..."

Wer sich anschickt, nicht nur schlechthin über Leitbilder der Technik, sondern speziell über deren Funktionen, Bedeutungen und Potentiale für den KI-Diskurs zu sprechen, gerät sehr schnell in Verlegenheit; auf die Frage nämlich, was KI eigentlich ist, erhält er, so Derek Partridge, von drei Praktikern mindestens vier verschiedene, und zwar definitive Antworten (Partridge 1986, S. 31). Doch damit nicht genug. Mehr noch als die Meinungsverschiedenheiten unter den KI-Spezialisten sorgen die interdisziplinären Kontroversen über die Bilanzen und Perspektiven des gesamten KI-Projektes für Irritation (Dreyfus 1989; Dreyfus/Dreyfus 1987; Winograd/Flores 1989;). Zugespitzt formuliert, schwankt der an KI interessierte Laie zunächst zwischen zwei Extremen: einerseits sieht er sich - noch Minskys, Simons oder Newells Prognosen im Ohr (Minsky 1967; .i.Simon/Newell 1958) - schon kurz vor der von Moravec prophezeiten "genetischen Wachablösung" (Moravec 1990, S. 13) auf der sich die menschliche Intelligenz für immer von ihrer Protein-"Sülze" (Moravec 1990, S. 163) lösen soll; andererseits ist er geneigt, derartige Szenarien schlicht unter der Rubrik Science Fiction abzubuchen, denn folgt man etwa Varela, dann stellen gegenwärtig noch die winzigsten Insekten und die simpelsten kognitiven Handlungen von Kleinkindern die KI-Forscher vor erhebliche Rätsel (Varela 1990, S. 57, 94). Eine pointierte Beschreibung des Forschungsfeldes lautet: "There's no business like show business - mit einer Ausnahme: KI." (Partridge 1986, S. 31) Der ironische und wohl auch etwas provokante Unterton ist zwar unüberhörbar, dennoch trifft dieser Vergleich unseres Erachtens einen, für den gesamten KI-Diskurs zentralen Punkt ziemlich genau, nämlich die anhaltende Faszination, die dieses Forschungsfeld auf Experten wie Laien ausübt. Inwiefern? Vertieft man sich als Außenstehender in den professionellen KI-Diskurs, dann sieht es so aus, als ob er sich einerseits um einen Kristallisationspunkt und andererseits um eine Leerstelle dreht.

1.2 Die Machbarkeit als diskursiver Kristallisationspunkt

Der Kristallisationspunkt betrifft die Frage nach der Machbarkeit der KI. Seitdem das KI-Projekt auf dem legendären Dartmouth-Seminar aus der Taufe gehoben wurde (Pape 1988, S. 446; Report 1992, S. 138), fehlt es nicht an Versuchen, immer wieder Bilanz zu ziehen. Die ursprünglichen Vorhersagen werden mit den schließlich eingetretenen Resultaten verglichen (Dreyfus 1989; Dreyfus/Dreyfus 1987; Coy/Bonsiepen 1987), und auch KI-Enthusiasten müssen einräumen, daß es da unübersehbare und zumeist erhebliche Diskrepanzen gibt (Report 1992, S. 66). Seit Jahrzehnten existiert eine mehr oder weniger systematische Differenz zwischen prognostizierter und realisierter Machbarkeit. Angesichts dieser Machbarkeits-Differenz sind nun folgende drei Tendenzen besonderes auffällig:

Erstens beeinträchtigt diese Differenz den Elan und Enthusiasmus der KI-Protagonisten kaum oder gar nicht. Das Sprichwort "if it works it's not AI" scheint viele KI-Forscher mehr zu erals zu entmutigen. Es liegt die Vermutung nahe, daß sich Machbarkeits-Differenzen und Forschungsanstrengungen nicht umgekehrt, sondern eher direkt proportional zueinander verhalten.

Zweitens scheint auch die staatliche und privatwirtschaftliche Förderungsbereitschaft durch die Machbarkeits-Differenz bisher nicht nachhaltig negativ beeinflusst zu werden. Glaubt man den Schätzungen, dann wurden bereits von 1985 bis 1987 allein in Deutschland rund 1 Mrd. DM für KI-Projekte ausgegeben (Schneider/Welsch 1988, S. 451). Und dies, obwohl sich abzeichnete, daß sich eine ganze Reihe optimistischer Vorhersagen - man denke etwa an die OVUM- und Hertz-Prognosen über die Umsatz-Entwicklung von KI-Produkten (Coy/Bonsiepen 1987, S. 83ff.) - nur sehr eingeschränkt beziehungsweise überhaupt nicht bewahrheiteten.

Drittens drehen sich viele Kontroversen zwischen Protagonisten und Kritikern der KI darum, ob bestimmte Forschungsziele überhaupt erreichbar oder nicht grundsätzlich zum Scheitern verurteilt sind. Der Streit um solche Fragen, ob sich Intelligenz auf kalkulierende Rationalität (Winograd/Flores 1989, S. 36ff., 157;; Report 1992, S. 28ff.), Repräsentationsbildung (Varela 1990, S. 37; Report 1992, S. 38ff.; Maturana/Varela 1987) oder Informationsverarbeitung (Diskurs-Protokoll II-1 1991, S. 9, 31; Maturana 1990, S. 15ff.; Report 1992, S. 34, 138) reduzieren läßt, welche Rolle Hintergrund- und Alltagswissen (Dreyfus 1989, S. 10f.; Dreyfus/Dreyfus 1987, S. 37ff) oder Emotionen und Motivationen (Report 1992, S. 26, 41; Dreyfus/Dreyfus 1987, S. 215ff.; Diskurs-Protokoll II-1 1991, S. 23, 31) spielen, wodurch sich Expertenwissen tatsächlich auszeichnet (Dreyfus/Dreyfus 1987, S. 54, 142; Coy/Bonsiepen

1987) oder welche Bedeutung der Körperlichkeit des Menschen zukommt (Diskurs-Protokoll II-1 1991, S. 12, 13, 25, 26, 29, 37, 38; Report 1992, S. 42; Maturana/Varela 1987), kreist vielfach um das Problem der generellen Machbarkeit der KI-Projekte.

Stimmt diese zugegebenerweise grobe Skizze in etwa, dann ergibt sich eine doch recht merkwürdige Situation, die sich zugespitzt wie folgt beschreiben ließe: Es gibt eine permanente Machbarkeits-Differenz mit einer gegenläufigen Motivationswirkung. Während sie die Promotoren und Sponsoren der KI-Forschung stets aufs Neue anspornt, diese Differenz zu schließen und die Machbarkeit ihrer Projekte unter Beweis zu stellen, fordert sie die Kritiker immer wieder dazu heraus, deren prinzipielle Nichtmachbarkeit nachzuweisen. In beiden Fällen kreisen die Argumentationen um die Machbarkeit - aber in entgegengesetzter Richtung und zuweilen auch aneinander vorbei. Beschränkt man sich auf diesen Kristallisationspunkt des KI-Diskurses, dann läßt sich die Gegensätzlichkeit der Expertenurteile nur bedingt, die anhaltende Faszination des KI-Projektes im Grunde nicht verstehen.

1.3 Die Wünschbarkeit als diskursive Leerstelle

Es ist deshalb nötig, sich etwas eingehender mit dem zweiten, weniger offensichtlichen Punkt zu beschäftigen, um den sich die Diskussionen drehen, nämlich die Leerstelle im professionellen KI-Diskurs. Diese Leerstelle betrifft die Frage nach der Wünschbarkeit der KI. Zwar wird auch immer wieder darüber gesprochen (Dreyfus/Dreyfus 1987, S. 269ff.; Report 1992, S. 138ff.), aber im Vergleich zur Frage der technischen Machbarkeit weit weniger, nicht so systematisch und eher am Rande. Es wäre allerdings verfehlt, deshalb die Bedeutung der Wünsche im KI-Diskurs gering zu veranschlagen. Im Gegenteil, denn es spricht manches dafür, da das KI-Projekt implizit - und insofern ist es dem show business durchaus ähnlich - gerade an die unerfüllten Wünsche, ungestillten Sehnsüchte und auch verdrängten Ängste vieler Menschen anknüpft.

Ganz so ungewöhnlich wie es zunächst scheint, ist diese Vermutung nicht, denn man täte der KI-Forschung sicher unrecht, wenn man ihre Anziehungskraft ausschließlich auf der naturbeziehungsweise ingenieurwissenschaftlichen Ebene verorten würde. Neben anderen sind es vor allem folgende vier Quellen, aus denen sich diese Anziehungskraft speist:

Erstens, die speziellen Denkzeuge. Hierzu zählen - um einmal eins der wohl faszinierendsten Beispiele aus dem Grenzbereich zwischen KI und Computeranimation herauszugreifen - die unter dem Stichwort "Cyberspace" oder "virtuelle Welten" laufenden Forschungen

(Rheingold 1992; Franke 1992). Die Simulation von molekularen Räumen, Gebäudeentwürfen oder Flugsituationen erweitert nicht nur das professionelle Arsenal der Chemiker, Architekten und Piloten. Wenn man nur einmal daran denkt, wie anschaulich und variantenreich sich städtebauliche Veränderungen zwischen allen Beteiligten cyberspacegestützt durchdiskutieren lassen, dann zeigt dies, welche neuen Möglichkeiten sich auch für die Vermittlung expertenkultureller und alltagsweltlicher Argumentationslogiken ergeben.

Zweitens, die universellen Erklärungsmodelle. Erinnerung sei hier nur an den physikalischen Reduktionismus (Diskurs-Protokoll II-1 1992, S. 18) und die Computeranalogie (Report 1992, S. 21ff.). Die Vorstellung, alle relevanten psychischen - und damit tendenziell auch sozialen - Geschehnisse ließen sich, vom Prinzip her, letztlich mit einer einzigen Theorie, nämlich der Physik - die überdies noch als Musterbeispiel der "hard science" (Foerster 1985, S. 17ff.), ja der Wissenschaften überhaupt gilt -, hinreichend erklären, ist ebenso verlockend wie die Annahme, das menschliche Gehirn und die digitale Rechenmaschine wären funktional äquivalent. Verlockend deshalb, weil sich in beiden Fällen nicht nur die Anzahl der konkurrierenden und oft genug auch widersprüchlichen Erklärungsmodelle, derer man sich bedienen muß, um Wissen über die Wirklichkeit zu gewinnen, reduziert, sondern auch, weil das so erworbene Wissen stringenter und fundamentaler, schlicht besser zu sein scheint.

Drittens, die alltagspraktische Gestaltungsmacht. Sie stößt, das müssen wohl alle Menschen im Laufe ihres Lebens schmerzlich feststellen, immer wieder an Grenzen. Eine dieser Grenzen liegt dort, wo sie auf die geistigen oder manuellen Fähigkeiten anderer Menschen zurückgreifen müssen, diese aber entweder nicht verfügbar oder nicht willens beziehungsweise in der Lage sind (Schleese 1992, S. 17f.), ihre Fähigkeiten bereitzustellen. Knowledge-Engineering und Robotik scheinen sich hier als Möglichkeiten anzubieten, traditionelle und bisher für unüberwindbar gehaltene Grenzen alltagspraktischer Gestaltungsmacht, wenn nicht zu beseitigen, so doch ein gutes Stück weiter hinauszuschieben.

Viertens, die emanzipatorischen Visionen. Hier sind es unter anderem zwei Vorstellungen, die eine besondere Anziehungskraft ausüben. Einmal ist es die Verheißung einer harmonischen und gerechten Welt, in der endlich eine reine Vernunft im Einklang mit den Gesetzen des Universums regieren kann, weil es ihr gelang, sich ein für alle mal von ihrem bisherigem Träger, dem homo sapiens, zu emanzipieren (Diskurs-Protokoll II-1 1991, S. 46, 49, 56; Report 1992, S. 71; Moravec 1990). Zum anderen ist es die Hoffnung, der Mensch könnte dem Tod entrinnen, indem es dem Bewußtsein gelingt, sich aus dem Gehäuse seines sterbenden Körpers zu befreien, um in ein ewiges Leben zu treten (Diskurs-Protokoll II-1 1991, S. 31/32).

Vergegenwärtigt man sich nun, daß die Wünsche nach speziellen Denkzeugen, universellen Erklärungsmodellen, alltagspraktischer Gestaltungsmacht und emanzipatorischen Visionen nicht nur so alt sind wie die Menschheit, sondern auch auf absehbare Zeit kaum aus den Menschen weichen werden, und bedenkt man ferner, daß heute, an der Schwelle zum 21. Jahrhundert, in den modernen Industriegesellschaften all jene sozialen Kräfte, die in der Vergangenheit eine Erfüllung dieser Wünsche verhießen - ob Religionen oder Wissenschaftsdisziplinen, ob Staat oder Familie, ob traditionelle oder neue soziale Bewegungen - offensichtlich an Ausstrahlungskraft und Einfluß verlieren (Zapf 1991), dann läßt sich das daraus resultierende Defizit schwerlich übersehen. Nun kann natürlich zu Recht gefragt werden, inwiefern dieses Defizit das KI-Projekt betrifft, ist doch immer wieder darauf verwiesen worden, daß KI-Protagonisten in ihrem Forschungshandeln als Rationalisten par excellence menschlichen Wünschen - insbesondere, wenn sie derart vage und diffus sind - eher gleichgültig oder reserviert gegenüberstehen, weil sie sich der kalkulierenden Vernunft entziehen. Daraus jedoch zu folgern, daß ihnen derartige Sehnsüchte fremd sind oder daß sie mit dem zuvor beschriebenen Defizit nicht rechnen und argumentieren würden, wäre voreilig, denn es gibt Grund zu der Annahme, daß das Wünschbare gerade dann, wenn es nicht direkt ausgesprochen wird, sondern mehr als unreflektierte Hintergrundgewißheit fungiert, die Wahrnehmungs-, Denk- und Argumentationslogik der KI-Forscher zwar diffizil, nichtsdestoweniger nachhaltig beeinflußt (Report 1992, S. 28ff., 38ff.).

2. Die Leitbild-Perspektive: Die Projektionen - Fusion, Kreislauf und Potential

Läßt man nun noch einmal das bisher zur Machbarkeit und Wünschbarkeit Gesagte Revue passieren, dann wird ein wenig besser verständlich, worauf sich die anhaltende Faszination des KI-Projektes gründet. Weder das Machbare noch das Wünschbare bilden für sich genommen das Fundament dieser Faszination: Das Machbare nicht, weil die dort erzielten Erfolge - gemessen an den Prognosen - vergleichsweise bescheiden, um nicht zu sagen niederschmetternd sind und das Wünschbare nicht, weil die unterschwelligsten Sehnsüchte - so sehr sie die Menschen auch gefangen nehmen mögen - längst ihre motivierende Wirkung verloren hätten, wenn sie sich nicht wenigstens partiell und ansatzweise, sozusagen keimhaft, erfüllen würden. Worauf also die anhaltende Faszination des KI-Projekts beruht, ist die Fusion oder Synthese zwischen Machbarkeits- und Wunschprojektionen, speziell die Tatsache, daß sich diese Fusion immer wieder selbst neu erzeugt.

Es gibt einen Kreislauf, der unseres Erachtens dem KI-Projekt insgesamt - und nicht nur ihm - zugrunde liegt: Wunsch- und Machbarkeitsprojektionen fusionieren in einer Vision, die über das durch sie geleitete Forschungshandeln bruchstückhaft in prototypischen technischen Lö-

sungen auskristallisiert, wodurch beide Projektionen an Stabilität gewinnen und neue Impulse bekommen, um erneut in denk- und handlungsleitenden Visionen zu verschmelzen, die wiederum zu neuen technischen Lösungen führen. Und genau dies ist jener Kreislauf, in dem sich die gestalterische Kraft technischer Leitbilder entfaltet. Leitbilder bündeln die Intuitionen und das (Erfahrungs-)Wissen der Menschen darüber, was ihnen einerseits als machbar und andererseits als wünschbar erscheint (Dierkes 1988, S. 54; Dierkes/Hoffmann/Marz 1992, S. 42). Es ist nicht zuletzt die Synthese von Machbarkeits- und Wunschprojektion, durch die sich technische Leitbilder auszeichnen. Insofern unterscheiden sie sich von konkreten einzelnen Forschungs- und Entwicklungszielen durch ihren vergleichsweise hohen Wunsch- und Phantasiegehalt, und von puren Illusionen durch ihre allgemein anerkannte - und sei es auch nur ansatzweise - technische Machbarkeit. Aus eben dieser Synthese, in der menschliche Träume, Visionen und Hoffnungen mit allgemein akzeptierten Wissensbeständen, Kalkülen und Prognosen verschmelzen, resultiert die Anziehungskraft und Faszination technischer Leitbilder wie auch ihr Einfluß auf den Prozeß der Technikentwicklung. Nicht auf ihren jeweils einzelnen Komponenten - dem Machbaren und/oder Wünschbaren -, sondern auf deren spezifischer Fusion gründet ihre Attraktivität und Stabilität. Das Machbare gilt als wünschbar und das Wünschbare als machbar. Man verfehlt deshalb die Funktionen, Bedeutungen und Potentiale technischer Leitbilder, wenn man sie nur auf die Dimension des technisch Machbaren herunterbuchstabiert. Dies um so mehr, als die orientierende, motivierende und formierende Wirkung von Leitbildern nicht zwangsläufig mit steigendem Wunsch- und Phantasiegehalt abnehmen muß.

Ausgehend von dem bisher Gesagten erschiene es uns sinnvoll, sich weit mehr als bisher auf jene Punkte zu konzentrieren, wo die Machbarkeits- und Wunschprojektionen der verschiedenen Akteursgruppen konkret verschmelzen. Eine solche Umorientierung auf einen leitbildzentrierten KI-Diskurs wäre unseres Erachtens unter anderem aus folgenden Gründen erfolgversprechend:

Erstens stößt man - wenn man einmal den KI-Diskurs aus der Leitbildperspektive betrachtet - auf eine Vielzahl solcher implizit vorhandener Leitbilder, die jedoch nur selten explizit zur Sprache gebracht und systematisch verhandelt werden. Viele Kontroversen wären nicht nur besser verständlich, sondern ließen sich auch präziser auf den Punkt und fallweise auch zu Ende bringen, wenn man sich zielgerichtet den konkurrierenden Leitbildern, die ihnen zugrunde liegen, zuwendet. Es sind, so unsere Annahme, wesentlich die unterschiedlichen technischen Leitbilder, die den KI-Diskurs implizit strukturieren und einen gut Teil der traditionellen Kontroversen immer wieder neu reproduzieren. Eine Leitbildzentrierung kann somit

dazu beitragen, die implizite Logik des KI-Diskurses transparent zu machen und zur Diskussion zu stellen.

Zweitens legen erste Untersuchungen zur Funktion und Bedeutung von Leitbildern in technikgenetischen Prozessen die Vermutung nahe, daß sie dort eine wichtige, wenn nicht gar zentrale Rolle spielen. (Dierkes/Marz 1991, 175ff.; Dierkes/Hoffmann/Marz 1992, 121ff.; Marcinkowski 1991, 9). Nicht zuletzt deshalb, weil sie für Gruppen und Individuen wahrnehmungs-, denk-, entscheidungs- und verhaltensleitend sowie für Organisationen und Institutionen - die ja zunehmend zu Trägern und Gestaltern der Forschungs- und Entwicklungsprozesse geworden sind (Dierkes 1989b; Dierkes 1990a) - kooperations-, koordinations- und kommunikationsleitend sind. Metaphorisch gesprochen handelt es sich bei Leitbildern um "Knotenpunkte" technikgenetischer Prozesse. Durch einen leitbildzentrierten KI-Diskurs kann es gelingen, die Reichweite und Folgenhaftigkeit der unterschiedlichen Forschungspfade besser in den Blick zu bekommen.

Drittens schließlich eröffnet sich über einen leitbildzentrierten KI-Diskurs nicht nur eine neue analytische Perspektive, sondern es erschließen sich auch konkrete Ansatzpunkte für eine prospektive Technikgestaltung. In einer professionsübergreifenden Diskussion um die den verschiedenen KI-Projekten zugrundeliegenden technischen Leitbilder wird nicht nur deutlicher erkennbar, welche Entwicklungsziele favorisiert - und welche folglich marginalisiert oder gänzlich ausgeblendet werden -, sondern es kann auch danach gefragt und darüber gestritten werden, welche Konsequenzen sich aus derartigen Ausblendungen (Hoffmann/Marz 1992; Schlese 1992, S. 22f.) beispielsweise für bestimmte soziale Gruppen ergeben und ob beziehungsweise inwieweit diese akzeptabel sind (Barben/Dierkes 1991, S. 218). Leitbild-Assessment wäre vielleicht der treffendste Begriff für dieses Herangehen, das darauf zielt, die Forschungs- und Entwicklungsprozesse über eine Gestaltung der technischen Leitbilder zu beeinflussen (Barben/Dierkes 1990, S. 443).

Die Notwendigkeit einer Umorientierung auf einen leitbildzentrierten KI-Diskurs zeigt sich vor allem dann, wenn man sich den Funktionen technischer Leitbilder, insbesondere ihrer Orientierungs-, Motivierungs- und Formierungskraft - und damit ihrer Folgenhaftigkeit - zuwendet. Bevor wir dies im folgenden stichpunktartig tun, gilt es jedoch zwei möglichen Erwartungshaltungen vorzubeugen: erstens stützen sich diese Überlegungen auf eine relativ schmale und KI-ferne empirische Basis (Dierkes/Hoffmann/Marz 1992, S. 60ff.; Marz/Dierkes 1993), tragen also in bezug auf das Forschungsfeld "Künstliche Intelligenz" eher heuristischen als analytischen Charakter; zweitens werden die Bezugspunkte zum KI-Projekt lediglich exemplarisch angedeutet und nicht systematisch entwickelt. Es bliebe

gesonderten Untersuchungen vorbehalten, die Tragfähigkeit dieser Ansätze in bezug auf das KI-Projekt im allgemeinen und KI-Diskurse im besonderen näher auszutesten und auszuloten.

3. Leitbild-Funktionen: Die Prozesse - Orientieren, Motivieren und Formieren

3.1 Das Orientieren oder der Forschungshorizont

In Fallstudien zu sehr unterschiedlichen technikgenetischen Feldern (Dierkes/Hoffman/Marz 1992, S. 60ff.; Marz/Dierkes 1993) zeigte sich immer wieder, daß technische Leitbilder keine bloßen Worthülsen sind, die über den Köpfen der Menschen schweben, sondern daß sie sehr wohl deren alltägliches Wahrnehmen, Denken, Entscheiden und Handeln bei der Gestaltung und Nutzung von Technik dauerhaft, nachhaltig und folgenschwer beeinflussen. Ein erster Grund dafür dürfte in der kognitiv-kommunikativen Orientierungsfunktion technischer Leitbilder zu suchen sein, die diese in dreierlei Hinsicht ausüben (Dierkes/Hoffmann/Marz 1992, S. 41ff.):

Erstens adaptieren Leitbilder die an der Produktion von technischem Wissen beteiligten Akteure permanent an künftige Kommunikations- und Kooperationsprozesse synchron vor, noch ehe sie in diesen Prozessen aufeinandertreffen. Die je verschiedenen persönlichen Wahrnehmungsmuster, die aus der Unterschiedlichkeit individueller Dispositionen, der Verschiedenartigkeit sozialer Positionen und aus der Spezifik der jeweiligen Profession resultieren, werden aufeinander vorabgestimmt. Diese Abstimmungsleistung ist keine Standardisierung oder Homogenisierung, in der die spezifischen Wahrnehmungsmuster eingeebnet und einander gleich gemacht werden. Es handelt sich hierbei vielmehr darum, daß alle den gleichen Wahrnehmungshorizont besitzen - bildlich gesprochen in dieselbe Richtung sehen -, was einschließt, daß andere Blickrichtungen gemeinsam ausgeschlossen werden.

Zweitens stellen technische Leitbilder eine Art funktionales Äquivalent dar. Forscher und Entwickler, Ingenieure und Wissenschaftler verfügen über professionsspezifische Regelsysteme und Entscheidungslogiken, mit deren Hilfe gemeinhin entschieden werden kann, was vernünftig, richtig und sinnvoll ist. Diese stellen für sie jedoch nicht nur ein Instrument dar, mit dem sie Probleme und Konflikte lösen, sondern stets auch einen Zwangsmechanismus, denn sie schaffen dialogische Fremd- und logische Selbstzwänge, die ständig ineinander umschlagen und sich wechselseitig stützen. Dort, wo neue technische Entwicklungspfade beschriftet werden und Vertreter unterschiedlicher Professionen miteinander in Kontakt kommen, würden die alltäglichen Kommunikations- und Kooperationsprozesse erschwert oder gar völlig zum Stillstand gebracht werden, wenn man sich dabei ausschließlich bestimmter ex-

pertenkultureller Regelsysteme und Entscheidungslogiken bediente. Leitbilder spielen hier die Rolle von funktionalen Äquivalenten, denn sie sind Zielpunkte, auf die sich die Diskurse ausrichten, beziehungsweise Orientierungslinien, entlang denen sie sich entwickeln können. Für jeden der Akteure - welcher speziellen Expertenkultur er auch angehören mag - "dreht" sich der Dialog um den gleichen Punkt. Grob zugespitzt: Es reicht nicht aus, den gleichen Wahrnehmungshorizont zu haben, man muß sich auch über das, was man aus den je verschiedenen expertenkulturellen Beobachtungsperspektiven wahrgenommen hat, untereinander verständigen können. Man denkt, diskutiert und entscheidet zwar auf unterschiedlichen Wegen, aber in die gleiche Richtung.

Drittens schließlich fungieren technische Leitbilder als problemorientierte Denkzeuge. Manche Probleme werden gedanklich leichter in einer begrifflichen, andere leichter in einer bildlichen Repräsentationsform bewältigt. Vieles spricht nun dafür, daß sich innovatives und kreatives Denken unter anderem dadurch auszeichnet, erstens die für jedes Problem adäquate Repräsentationsform zu finden - also jene Form, in der es die einfachste und unkomplizierteste Gestalt annimmt - und zweitens diese beiden Repräsentationsformen wechselseitig ineinander zu transformieren können (Klix 1983, 268ff.). Leitbildgestützt lassen sich diese beiden Aufgaben leichter bewältigen. Erstens dienen technische Leitbilder selbst als bildlicher Repräsentationsrahmen, in dem sich neue, ungewohnte Probleme denken lassen und zweitens können sie der ständigen Umwandlung von begrifflichen in bildliche Repräsentationsformen eine Richtung geben. Die Menschen, die sich in ihrem Denken von einem Leitbild leiten lassen, tun dies also, indem sie vermittels des Bildes und in dem Bild denken.

3.2 Das Motivieren oder die Forscher

Doch die Funktion technischer Leitbilder erschöpft sich nicht in der kognitiv-kommunikativen Orientierung. Es sind - bis jetzt jedenfalls - nicht technische Artefakte, sondern Menschen, die einen bestimmten Forschungshorizont vor Augen haben und sich abmühen, ihm Schritt um Schritt näher zu kommen. Und bei diesen Bemühungen handeln sie nicht nur als bloße "Wahrnehmungsapparate" oder "Denkmaschinen", sondern als Persönlichkeiten, deren Leistungsfähigkeit motivationsabhängig ist. In bezug auf die Motivation der Forscher sind technische Leitbilder mindestens in zweifacher Hinsicht von Bedeutung (Dierkes/Hoffmann/Marz 1992, S. 54f.):

Erstens aktivieren sie die emotionalen, volitiven und affektiven Potentiale der Menschen. Sie mobilisieren damit nicht nur den Kopf, sondern die ganze Persönlichkeit. Ein technisches Leitbild residiert nicht nur in den Hirnen der Forscher, es berührt auch - man gestatte diese

unwissenschaftliche Formulierung - ihre Herzen. Es besteht kein Anlaß, die Bedeutung dieser Leitbildfunktion gering zu veranschlagen. Im Gegenteil, ob und inwieweit es gelingt, einen neuen und deshalb zwangsläufig auch mit vielen Unsicherheiten und Unwägbarkeiten behafteten Entwicklungspfad erfolgreich zu beschreiten, dürfte zu einem nicht gerade geringen Teil vom Grad der persönlichen Involviertheit der Menschen abhängen, davon, wie viele Akteure wie tief "mit allen Fasern" in diesen Prozeß eingebunden sind.

Zweitens setzt ein erfolgreiches Zusammenwirken verschiedener Wissenskulturen in einem Forschungsfeld stets voraus, daß ihre Vertreter tagtäglich Kooperationsleistungen erbringen. Die Formen, in denen dies geschieht, können sich jedoch erheblich voneinander unterscheiden. Nicht zuletzt dadurch, ob die jeweiligen Kooperationsleistungen den Akteuren nur von außen durch soziale Fremdwänge abgepreßt und abgetrotzt werden, oder ob sie auch von innen, durch die psychischen Selbstzwänge der Menschen, durch ihren Willen und ihre Fähigkeit zur Kooperation zustande kommen. Technische Leitbilder bewirken nun eine vergleichsweise kosten- und verlustarme alltägliche Kooperation. Die Arbeit an neuen technischen Entwicklungen kann nicht reibungs- und konfliktlos verlaufen, sondern wird zwangsläufig zu psychischen Spannungen in und sozialen Spannungen zwischen den daran beteiligten Menschen führen. Leitbilder vermögen diese unvermeidlich auftretenden Spannungen zu dämpfen. Sie verbinden Menschen, die sonst nichts aneinander bindet; Menschen, die nicht nur verschiedenen sozialen Milieus, sondern auch unterschiedlichen (Sub)kulturen angehören, deren Wahrnehmen, Denken, Entscheiden und Verhalten deshalb unter Umständen nicht nur einfach abweichenden, sondern geradezu gegensätzlichen Orientierungen folgen kann. Leitbilder motivieren die Forscher - über viele Konflikte und Rückschläge hinweg - zur Weiter- und Zusammenarbeit.

Auch wenn in bezug auf eine ganze Reihe von KI-Techniken, denen zunächst eine spektakuläre Karriere vorausgesagt wurde - man denke etwa an Expertensysteme oder Heimroboter (Coy/Bonsiepen 1987, S. 83ff.) - zwischenzeitlich Ernüchterung Raum gegriffen hat und gegenwärtig nicht zu befürchten ist, von den Erfolgen der KI-Forschung auf breiter Front alltagspraktisch überrollt zu werden, besteht kein Anlaß, die Reichweite und Folgenhaftigkeit von KI-Visionen zu unterschätzen (Coy/Bonsiepen 1987, S. 15; Die Mitbestimmung, Nr. 9/88; Report 1992, S. 138ff.). Dies wird deutlich, wenn man sich der Formierungsfunktion technischer Leitbilder zuwendet.

3.3 Das Formieren oder die Forschungspfade

Wenn über technische Leitbilder eine Grund- oder Vorformatierung (Dierkes/Marz 1992a, S. 232) der Wahrnehmungs-, Denk-, Entscheidungs- und Handlungsmuster der Akteure techniken-genetischer Prozesse bewirkt und so ein bestimmter Forschungshorizont abgesteckt wird, dann stellt sich die Frage, inwiefern dadurch die verschiedenen Forschungspfade formiert und die auf diesen Pfaden entwickelten Techniken geprägt werden. Bringt man diese Fragestellung auf den Punkt, geht es darum, ob und inwieweit sich Leitbilder in die konkreten Formen und Gestalten der technischen Artefakte, Technologien oder großtechnischen Systeme einschreiben. Natürlich wäre es absurd anzunehmen, neue Techniken wären in einem Leitbild bis ins letzte Detail vorprogrammiert. Technische Leitbilder sind keine Konstruktionspläne oder Aufbauanleitungen. Wenn Leitbilder tatsächlich neue Techniken prägen, müssen sie sich zwar in deren Funktionsprinzipien und stofflich-gegenständlichen Formen einschreiben, dürften diesen jedoch nicht ohne weiteres abzulesen sein. Wie man einem Menschen nicht ansehen kann, welche spezifische Genstruktur ihn wie weit geprägt hat, wird man Techniken nicht ohne weiteres ihre Leitbildprägung ansehen. Aus der Tatsache, daß die Leitbildprägung nicht auf der Hand liegt, umstandslos den Schluß zu ziehen, daß es sie nicht gibt, wäre allerdings kurzschlüssig und voreilig. Ein möglicher Weg, die Frage, wie Leitbilder Forschungspfade formieren und Techniken prägen, der Beantwortung näher zu bringen, besteht darin, zwischen feldgenerierenden und pfadselektierenden Leitbildern zu unterscheiden. Ausgehend von einer solchen Unterscheidung ergeben sich dann Fragen wie die folgenden: Gibt es technische Leitbilder, durch die Forschungsfelder generiert werden und solche, die dann in diesen Feldern bestimmte Entwicklungspfade präferieren, andere folglich marginalisieren oder blockieren? Und, wenn solche Leitbilder auszumachen sind, besteht ein Zusammenhang zwischen feldgenerierenden und pfadselektierenden Leitbildern? Schließlich: Läßt sich diese doppelte Leitbildprägung an den technischen Systemen zeigen? Ausgehend von grundsätzlichen Überlegungen (Dierkes 1990b; Dierkes/Hoffmann 1992) sind wir diesen Fragen auch speziell im Rahmen unseres Forschungsprojekts "Technikgenese in organisatorischen Kontexten" (Dierkes 1988; Dierkes 1989a) zunächst an Hand einer technikgeschichtlichen Detailstudie nachgegangen und haben erste vorläufige Teilantworten zur Diskussion gestellt (Marz/Dierkes 1993), die hier allerdings nicht referiert werden können. Nun ist es immer problematisch, aus Einzelfällen Schlüsse abzuleiten, dennoch liegt die Vermutung nahe, daß die Dominanz bestimmter feldgenerierender und pfadselektierender Leitbilder einen nachhaltigen Einfluß auf die Formierung der Techniken und die sich daraus ergebenden Nutzungskontexte ausübt (Dierkes/Marz 1992b).

Wenn das so ist, wäre ein Leitbild-Assessment ein möglicher und erfolgversprechender Weg für eine prospektive Technikfolgenregulierung. Einerseits zwingt dies nämlich die Vertreter der entsprechenden Expertenkulturen, ihre nicht gerade selten schwer durch- und überschaubaren Argumentationsmuster auf einen einsichtigen und beurteilbaren Kern zu bringen. Andererseits räumt dies vielen Menschen, die nicht diesen Expertenkulturen angehören und die deren exklusiven Begründungslogiken oft rat- und hilflos gegenüberstehen, sehr wohl jedoch über kurz oder lang von den Ergebnissen der expertenkulturellen Arbeit betroffen sind, die Chance ein, über Leitbild-Diskurse Einfluß auf technische Entwicklungslinien zu nehmen. Ein solches Leitbild-Assessment, also der Gedanke, Leitbilder zu gestalten, um so Prozesse der Technikentstehung zu beeinflussen und die Interessen unterschiedlicher gesellschaftlicher Gruppen rechtzeitig zu berücksichtigen und aufeinander abzustimmen, ist ja gar nicht so spektakulär, wie es auf den ersten Blick den Anschein hat. Was hat denn beispielsweise die Enquête-Kommission "zukünftige Kernenergie-Politik" des Deutschen Bundestages anderes getan, als in Pfadszenarien verschiedene energiepolitische Leitbilder zu entwickeln und diese dann auf ihre möglichen Folgen hin abzuschätzen? Gleiches gilt für die Szenarien zur Entwicklung der landwirtschaftlichen Produktionsweisen, die von der Enquête-Kommission "Technikfolgenabschätzung" auf ihre unterschiedlichen Konsequenzen hin überprüft wurden. Und, so könnte weiter gefragt werden, was tun denn eigentlich Ethik-Kommissionen etwa im Bereich der Gentechnologie anderes, als Leitbilder möglicher Entwicklungspfade dieser Technik aufzuspüren und auf ihre verschiedenen Folgen hin zu untersuchen? Technikleitbilder können also - in großen Bandbreiten wenigstens - auf ihre möglichen Folgen hin untersucht werden. Es können darüber hinaus ganz andere oder auch nur leicht veränderte Leitbilder für die Entwicklung eines Technikfeldes in die öffentliche Diskussion eingebracht und als Alternativen mit den Entscheidungsträgern der Technikentwicklung in diesem Feld diskutiert werden.

Sicher wäre es vermessen anzunehmen, daß sich über Leitbild-Diskurse sämtliche nicht intendierten Sekundär- und Tertiärfolgen bestimmter technischer Entwicklungspfade exakt vorhersehen ließen. Unnützlich indes wären solche Diskurse gewiß nicht, könnten sie doch dazu beitragen, das Formierungs- und Prägepotential technischer Leitbilder etwas genauer in den Blick und vielleicht auch ein wenig besser in den Griff zu bekommen (Dierkes/Hoffmann/Marz 1992, S. 153ff.).

4. Eine Chance: Leitbildforschung im KI-Diskurs

Gerade für das facettenreiche Forschungsfeld der "Künstlichen Intelligenz" ist dies unseres Erachtens von besonderer Bedeutung, denn mit Ausnahme der Bio- und Gentechnologie (i. Barben/Dierkes/Marz 1992;) dürfte es schwerfallen, wissenschaftlich-technische Felder zu finden, in denen Machbarkeits- und Wunschprojektionen derart faszinierend verschmelzen und deren Projekte eine so starke Orientierungs-, Motivierungs- und Formierungsfunktion ausüben. Dafür, die KI-Forschungen aus der Leitbildperspektive zu diskutieren, spricht auch, daß sich dadurch die Basis öffentlicher Diskussionen über dieses Technikfeld wesentlich erweitern würde. Ausgehend davon wäre zu überlegen, ob nicht gerade der vom Bundesministerium für Forschung und Technologie initiierte und vom VDI getragene Diskurs zum "Menschenbild in der Künstlichen Intelligenz" für eine leitbildzentrierte Untersuchungsperspektive mit Gewinn ausgewertet werden könnte. Aufbauend auf konzeptionellen Ansätzen zur Rolle von Leitbildern in der Technikgenese und bei der Kommunikation zwischen unterschiedlichen Wissenskulturen (Dierkes/Marz/Hoffmann 1992, S. 27ff.), empirischen Fallstudien (Dierkes/Marz/Hoffmann 1992, S. 59ff.) sowie Überlegungen zur Leitbildperspektive (Hoffmann/Marz 1992), Leitbildprägung und Leitbildgestaltung (Marz/Dierkes 1993) wäre dabei zu prüfen, ob sich die den Menschenbild-Diskurs formierenden und in ihm mehr oder weniger deutlich aufscheinenden Leitbilder klar herausarbeiten lassen. Davon ausgehend könnte dann deren Orientierungs- und Formierungskraft empirisch untersucht und an Hand ausgewählter Beispiele plausibel gemacht werden. Aus einer solchen grundsätzlichen Zielstellung würden sich dann folgende vier zentrale Fragestellungen für eine konkrete Diskursanalyse ergeben:

1. Lassen sich in den Diskursmaterialien zum "Menschenbild in der Künstlichen Intelligenz" überhaupt Leitbilder auffinden, und um welche handelt es sich dabei? (Die Leitbildrekonstruktion).
2. Spielen diese Leitbilder eine eher singuläre Rolle oder prägen sie den Verlauf und die Struktur des Menschenbild-Diskurses wesentlich, und wenn ja, auf welche Art und Weise? (Die Diskursformierung).
3. Ergeben sich aus dieser, durch Leitbilder bewirkten Diskursformierung Grenzen für die Technikfolgenabschätzung zur KI-Forschung, und worin bestehen sie? (Die Diskursgrenzen).

4. Gibt es konkrete konzeptionelle Ansatzpunkte, um diese Diskursgrenzen zu überwinden? (Die Diskursorientierung).

Angesichts der wechselvollen Geschichte und der nach wie vor kontrovers diskutierten Zukunftsperspektiven des KI-Projekts könnte eine solche leitbildzentrierte Diskursanalyse - so ist jedenfalls zu hoffen - für Laien, Protagonisten und Kritiker der "Künstlichen Intelligenz" gleichermaßen von Interesse sein.

5. Literatur

- Barben, D./Dierkes, M. (1990): Un-Sicherheiten im Streit um Sicherheit - Zur Relevanz der Kontroversen um die Regulierung technischer Risiken. In: Sarcinelli, U. (Hrsg.): Demokratische Streitkultur. Theoretische Grundpositionen und Handlungsalternativen in Politikfeldern. S. 422ff. Opladen
- Barben, D./Dierkes, M. (1991): Wirtschaftsethik, Unternehmenskultur und Technikfolgenabschätzung - Orientierungsgrundlagen für die tägliche Praxis? In: Dierkes, M./Zimmermann, K. (Hrsg.): Ethik und Geschäft. Dimensionen und Grenzen unternehmerischer Verantwortung. S. 205ff. Frankfurt a. M.
- Barben, D./Dierkes, M./Marz, L. (1992): Guiding Visions - Their Role in the Public Discourse and Development of Biotechnology. (im Druck)
- Coy, W./Bonsiepen, L. (1987): Erfahrung und Berechnung. Kritik der Expertensystemtechnik. Heidelberg et al. Die Mitbestimmung. Monatszeitschrift der Hans-Böckler-Stiftung, Nr. 9/88
- Dierkes, M. (1988): Organisationskultur und Leitbilder als Einflußfaktoren der Technikgenese. Thesen zur Strukturierung eines Forschungsfeldes. In: Verbund Sozialwissenschaftliche Technikforschung. Mitteilungen, Heft 3. S. 49ff. München
- Dierkes, M. (1989a): Technikgenese in organisatorischen Kontexten. Neue Entwicklungslinien sozialwissenschaftlicher Technikforschung. WZB-dp FS II 89-104. Berlin
- Dierkes, M. (1989b): Technikfolgenabschätzung in Unternehmen: Notwendigkeit, Möglichkeiten und Grenzen. In: Biervert, B./Dierkes, M. (Hrsg.): Informations- und Kommunikationstechniken im Dienstleistungssektor. Rationalisierung oder neue Qualität? S. 59ff. Wiesbaden
- Dierkes, M. (1990a): Ist Technikentwicklung steuerbar? In: Lenk, K. (Hrsg.): Programme zur Technikentwicklung - Technikentwicklung nach Programm? Werkstattbericht der Gesellschaft für Programmforschung in der öffentlichen Verwaltung e. V. S. 5ff. München
- Dierkes, M. (1990b): Technische Entwicklung als sozialer Prozeß. In: Naturwissenschaft, Heft 5/1990. S. 214ff.
- Dierkes, M./Hoffmann, U. (Hrsg.) (1992): New Technology at the Outset. Social Forces in the Shaping of Technological Innovations. Frankfurt/New York
- Dierkes, M./Hoffman, U./Marz, L. (1992): Leitbild und Technik. Zur Entstehung und Steuerung technischer Innovationen. Berlin
- Dierkes, M./Marz, L. (1991): Technikakzeptanz, Technikfolgen und Technikgenese. Zur Weiterentwicklung konzeptioneller Grundlagen der sozialwissenschaftlichen Technikforschung. In: Jaufmann, D./Kistler, E. (Hrsg.): Einstellungen zum technischen Fortschritt. Technikakzeptanz im nationalen und internationalen Vergleich. S. 157ff. Frankfurt/New York

- Dierkes, M./Marz, L. (1992a): Umweltorientierung als Teil der Unternehmenskultur. In: Steger, U. (Hrsg.): Handbuch des Umweltmanagements. Anforderungs- und Leistungsprofile von Unternehmen und Gesellschaft. S. 223ff. München
- Dierkes, M./Marz, L. (1992b): Technikleitbilder und Anwenderinteressen. Ein Beitrag zum vorbeugenden Verbraucherschutz. In: Verbraucherpolitische Hefte Nr. 15 (Dezember 1992) S. 39ff.
- Diskurs-Protokoll II-1. Workshop des VDI-Ausschusses "Künstliche Intelligenz" am 2. und 3. November 1990 in Berlin im Rahmen des Diskursvorhabens "Das Menschenbild in der Künstlichen Intelligenz" der VDI-Hauptgruppe "Der Ingenieur in Beruf und Gesellschaft". Berlin 1991
- Dreyfus, H. L. (1989): Was Computer nicht können. Die Grenzen künstlicher Intelligenz. Frankfurt/a. M.
- Dreyfus, H. L./Dreyfus, S. E. (1987): Künstliche Intelligenz. Von den Grenzen der Denkmaschine und dem Wert der Intuition. Hamburg
- Foerster, H. v. (1985): Sicht und Einsicht. Versuche zu einer operativen Erkenntnistheorie. Braunschweig/Wiesbaden
- Franke, H. W. (1992): Welten aus Bits und Bytes. In: VDI Nachrichtenmagazin, Heft 9/1992. S. 13ff.
- Hoffmann, U./Marz, L. (1992): Leitbildperspektiven. Technische Innovationen zwischen Vorstellung und Wirklichkeit. In: Burmeister, K./Steinmüller, K. (Hrsg.): Streifzüge ins übermorgen. Science Fiction und Zukunftsforschung. Weinheim/Basel
- Klix, F. (1983): Erwachendes Denken. Eine Entwicklungsgeschichte der menschlichen Intelligenz. Berlin (Ost)
- Marcinkowski, F. (1991): Technikkommunikation und geplanter Gesellschaftswandel. Theoretische Bezüge und empirische Fragestellungen eines Forschungsprogramms. Duisburger Materialien zur Politik- und Verwaltungswissenschaft Nr. 13. Universität- GH-Duisburg
- Marz, L./Dierkes, M. (1993): Leitbildprägung und Leitbildgestaltung. Zum Beitrag der Technikgenese-Forschung für eine prospektive Technikfolgen-Regulierung. In: Verbund Sozialwissenschaftliche Technikforschung. Mitteilungen Heft 10/1993. Technik-Gestalten-Lernen. Beiträge aus der Arbeit des Verbundes. Bremen. S. 95ff.
- Maturana, H. R. (1990): Gespräch mit Volker Riegas und Christian Vetter. In: Riegas, V./Vetter, Ch. (Hrsg.): Zur Biologie der Kognition. Frankfurt/a. M.
- 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/München/Wiesbaden
- Minsky, M. (1967): Computation: Finite and Infinite Machines. Englewood Cliffs
- Moravec, H. (1990): Mind Children. Der Wettlauf zwischen menschlicher und künstlicher Intelligenz. Hamburg
- Pape, M. (1988): Von der Rechenmaschine zur Künstlichen Intelligenz. In: Die Mitbestimmung, Heft 9/1988. S. 445ff.
- Partridge, D. (1986): Will AI lead to a Super Software Crisis. In: Gill, K. (Hrsg.): Artificial Intelligence for Society. S. 31ff. New York et al.
- Rheingold, H. (1992): Virtuelle Welten. Reisen im Cyberspace. Hamburg
- Schleese, M. (1992): Die KI ist tot. Es lebe die KI!. Vortrag, gehalten auf der Tagung der Sektion "Wissenschafts- und Technikforschung" der DGS in Zusammenarbeit mit der Sektion "KI und Gesellschaft" der Gesellschaft für Informatik in Bielefeld am 22. Mai 1992. (unveröffentlicht. Manuskript)

- Schneider, R./Welsch, J. (1988): Künstliche Intelligenz: Demokratisierung des Wissens oder Vorbote eines neuen Taylorismus? In: Die Mitbestimmung, Heft 9/1988. S. 449ff.
- Simon, H./Newell, A. (1958): Heuristic Problem Solving: The Next Advance in Operations Research. In: Operations Research 6/1958. S. 6ff.
- Varela, F. J. (1990): Kognitionswissenschaft - Kognitionstechnik. Eine Skizze aktueller Perspektiven. Frankfurt/a. M.
- VDI-Hauptgruppe "Der Ingenieur in Beruf und Gesellschaft" (Hrsg.) (1992): Künstliche Intelligenz. Leitvorstellungen und Verantwortbarkeit. VDI Report 17. Düsseldorf
- Winograd, T./Flores, F. (1989): Erkenntnis. Maschinen. Verstehen. Zur Neugestaltung von Computersystemen. Berlin
- Zapf, H. (Hrsg.) (1991): Die Modernisierung moderner Gesellschaften. Verhandlungen des 25. Deutschen Soziologentages in Frankfurt am Main 1990. Frankfurt/New York

Zukunftsvisionen der KI: Wünschbare und realistische Wirkungen auf die Lebenswelt*

Prof. Dr. I. Wachsmuth

Mit den hier vorgetragenen Ausführungen berichte ich über die Tätigkeit derjenigen Arbeitsgruppe in unserem Diskursprojekt, die sich mit möglichen Auswirkungen zukünftiger KI-Systeme auf die Gesellschaft auseinandersetzt.¹ Ging es in einem Teil des Diskurses vor allem um die gegenwärtigen Entwicklungen der KI und um ihre kritische Reflexion, soll der Ausgangspunkt für diesen Beitrag die Frage sein: Was ist eigentlich für zukünftige Entwicklungen wünschbar, und sind nicht technische Innovationen nicht nur vom Standpunkt des Denkbaren, sondern auch des Wünschbaren bzw. des Gewünschten zu diskutieren? Die Vorhersage zukünftiger Entwicklungen und ihrer Auswirkungen ist schwierig, wenn nicht unmöglich. Wenn man sich aber überlegt, aus welchen Richtungen kommen denn Wünsche, was sind Leitvorstellungen und Motivationen der Beteiligten, dann kann man vielleicht eher erkennen, wohin die Entwicklungen laufen könnten.

Eine weitere Bemerkung will ich vorausschicken: Wir diskutieren hier zu einem Teil fiktive Szenarien, von denen man bei realistischer Einschätzung sagen möchte, "so kommt das alles sowieso nicht". Allerdings erkennt man doch in dem, was in den Medien über Künstliche Intelligenz geschrieben wird, daß ein großer Bedarf besteht, sich mit solchen Zukunftsvisionen – die ja in vielen Fällen mit bereits begonnenen oder geplanten Forschungs- und Entwicklungsarbeiten entstehen – auseinanderzusetzen. Selbst wenn wir also von einem Insiderstandpunkt zu erkennen glauben, daß manche der großen Entwürfe nicht oder nur mit großen Abstrichen erfolgreich sein werden, erscheint es mir sinnvoll zu fragen, wie wäre denn eigentlich eine Gesellschaft mit "intelligenten" Maschinen, um sich anhand der Implikationen des Denkbaren mit dem Wünschbaren zu befassen. Mein wiederholt vertretener Standpunkt ist dabei, daß Künstliche Intelligenz eine besondere Herausforderung für die Zukunftsgestaltung darstellt und daß "Zukunft" keine passive Angelegenheit sein sollte, in die man sich treiben

* Transkript des Vortrags.

¹ G. Görz, A. Kremeier, H. Röpke, P. Schreiber, G. Strube, I. Wachsmuth, M. Wilker.

läßt, sondern daß die informationstechnische Zukunft mit KI aktiv zu gestalten ist. Denn die Richtungen, in die sich die Zukunft durch Künstliche Intelligenz entwickeln könnte, sind nicht völlig zufällig, sondern sie werden geleitet von mehr oder weniger expliziten Spekulationen der an der Technikentwicklung Beteiligten und ihrer Auftraggeber in Politik – auch Militär – und Wirtschaft.

Ich möchte nun zunächst einige Statements aus unserem Zukunftsdiskurs aufgreifen. Zuerst einmal hatten wir Schwierigkeiten, Künstliche Intelligenz von entwickelter Informationstechnik abzugrenzen. Und dies ist vielleicht auch schwer zu trennen, denn Techniken, die im Gebiet der KI entstanden sind, sind in vielen Fällen Ausgangspunkt informationstechnischer Innovationen gewesen (beispielsweise objektorientierte Programmiermethoden, die u.a. eine Basis heutiger Benutzungsoberflächen bilden). Wegen dieser grundsätzlichen Schwierigkeiten haben wir für unsere gegenwärtigen Überlegungen die folgende Arbeitshypothese an den Ausgang gestellt: Unter "entwickelter Künstlicher Intelligenz" wollen wir Systeme verstehen, die über kognitive Komponenten verfügen, die also beispielsweise Sensoren für optische und akustische Wahrnehmung haben, die solche Sensordaten verarbeiten können, die zu Schlußfolgerungen fähig sind, die heuristische Problemlösestrategien entwickeln und die derartige Fähigkeiten insgesamt in einem System integrieren. Wir beziehen uns also auf das, was man heute mit dem Stichwort "autonome Systeme" belegt, Systeme, die – mit Abstrichen – ähnlich wie der Mensch mit einer Außenwelt kommunizieren und interagieren.

Mit diesem Ausgangspunkt haben wir dann gestaffelte Entwicklungsstadien betrachtet – fiktive Stadien mit allen Vorbehalten über ihr tatsächliches Eintreten – und uns gefragt, angenommen, es käme so, was wären die Implikationen? In unserem Diskurs wollen wir also keine Aussagen darüber machen, was kommt, sondern die Frage diskutieren, "was wäre wenn?".

- Für das erste Stadium haben wir angenommen, daß Maschinen, die von Menschen entsprechend programmiert worden sind, intellektuelle Routinetätigkeiten übernehmen können. Dazu gehören einerseits klassische Expertensysteme, wie sie beispielsweise in der technischen Diagnostik bereits eingesetzt werden, des weiteren etwa maschinelle Systeme, die umfangreiche Meßdaten intelligent auswerten, indem sie beispielsweise Regularitäten abstrahieren; auch hierfür hat sich der Einsatz gegenwärtiger Techniken bereits als tragfähig und nützlich erwiesen.
- Für das zweite Stadium – für uns denkbar in den nächsten dreißig Jahren – galt folgender Gedanke: Wenn gegenwärtig projektierte Forschungsarbeiten erfolgreich sind, was wäre

dann als neue Qualität zu erwarten? Hierfür haben wir angenommen, daß Maschinen in größerem Umfang selbständig aus Beobachtungen, die sie vermöge ihrer Wahrnehmungskomponenten in ihrem Gegenstandsbereich anstellen, Wissen ansammeln; sie können auch unscharfes Wissen verwerten und eigene Fehler diagnostizieren. Dies ist sicherlich heute kaum auch nur ansatzweise möglich; es ist aber in Projektplänen derartiges vorgesehen.

- Für das dritte Stadium – über dreißig Jahre hinaus extrapoliert – haben wir spekulativere Zukunftsvisionen einbezogen, die annehmen, daß Maschinen nur noch teilweise determiniert sind, daß sie Situationen, mit denen sie konfrontiert sind, absichtsvoll verbessern und umgestalten können, indem sie aus sich selbst als Reaktion auf die Einwirkung der Umwelt kreatives Verhalten zeigen (etwa durch Strukturbildung und Selbstorganisation), daß sie Spontaneität, Zielstrebigkeit und Sensibilität besitzen können.

Ich betone noch einmal, wir haben diese hypothetischen Stadien einer entwickelten KI – die durchaus abgeleitet sind aus gegenwärtig diskutierten Zukunftsvorstellungen – als ein fiktives Szenario angenommen, und wir haben die damit verbundenen Entwicklungen nicht in Erwartung ihres wahrscheinlichen Eintretens als vielmehr hinsichtlich ihrer Implikationen und ihrer Wünschbarkeit betrachtet. Auch wenn die Möglichkeit der im dritten Stadium angesprochenen Entwicklungen mit Skepsis betrachtet wurde, darf nicht übersehen werden, daß derartige Zukunftsvisionen durchaus Einfluß auf Entscheidungsträger, etwa hinsichtlich der Finanzierung von Forschungs- und Entwicklungsprogrammen nehmen können. Der Punkt, um den es mir hier geht und der unsere spekulativen Überlegungen legitimiert, ist der folgende: Wenn aufgrund von projektierten Forschungs- und Entwicklungsarbeiten in breitem Umfang auf entsprechende Ziele hin gearbeitet wird, entstehen verwertete Teilergebnisse, und dann ist die Zukunft ein Stück weit anders, als wäre sie nicht von einer solchen Vision "gezogen" worden. Beispiele wie SDI haben gezeigt, daß bestimmte Leitvorstellungen kaum zurücknehmbare Auswirkungen haben, auch wenn die Visionen nicht im vollen Umfang erreichbar sind. Ihr Einfluß auf mögliche Zukunftsentwicklungen darf also nicht unterschätzt werden.

Die Fragen, mit denen wir uns in diesem Szenario auseinandergesetzt haben, haben sich zunächst einmal vornehmlich auf mögliche Veränderungen der Arbeitswelt bezogen; also etwa: Wird eine umfangreiche Neubewertung gegenwärtiger Qualifikationen eintreten? Welche Arbeit bleibt dem Menschen vorbehalten und wie wird sie dann bewertet? Verfällt beispielsweise ein Teil der Qualifikation des Ingenieurs? Werden sich grundsätzlich neue Strukturen in der Arbeitswelt entwickeln? Bilden wir noch zeitgemäß aus?

Aus unserem vorgelegten Bericht² möchte ich dazu die folgenden Überlegungen herausgreifen: Die technologische Weiterentwicklung kommt nicht von heute auf morgen über uns, pointiert ausgedrückt wird sie nicht über eine mehr oder weniger wehrlose Sozialgemeinschaft quasi "ausgegossen", sondern sie findet auf dem Umweg über die oft mühsame und langwierige Produktentwicklung Eingang in gesellschaftliche Wirkungszusammenhänge – nach Prinzipien der Markterschließung, Akzeptanz und der allmählichen Anpassung von Arbeitsorganisation und Arbeitsteilung. Soziale Systeme haben seit der Industriellen Revolution weitgehende Aufnahmefähigkeit für technologischen Fortschritt bewiesen. In vielen Fällen werden KI-Produkte als verbesserte Substitute bereits in Gebrauch befindlicher Artefakte in Erscheinung treten, wobei die Tatsache, daß ein Produkt nach seinem *Gebrauchswert* und kaum nach der ermöglichenden Technik bewertet wird, es fraglich erscheinen läßt, ob man darin verwandte KI-Techniken überhaupt als solche wahrnehmen wird.

Im weiteren Verlauf des Diskurses ist dann der Eindruck gewachsen, daß sich durch ein Vordringen von KI-Technologie in den Freizeitbereich ungleich umfangreichere Implikationen für die Alltags- und Lebenswelt ergeben könnten, und das soll im folgenden das Thema meiner Ausführungen sein. Um etwas konkreter auf das mögliche Vordringen von KI-Technologie in die Lebenswelt einzugehen, möchte ich einige Passagen aus einer Electronic Mail-Newsgroup³, die sich kürzlich mit diesem Thema befaßt hat, präsentieren. Dort ist unter der Frage "*AI in the year 2025*" – "Künstliche Intelligenz im Jahre 2025" – eine breitgestreute Umfrage überwiegend auf dem amerikanischen Kontinent gelaufen, die von ähnlichen Gedanken wie unser Vorgehen getragen war: "Was stellt Ihr Euch denn vor, das in dreißig Jahren ist?".

Von den zusammengestellten 35 Antworten, die auch Wünsche der Menschen erkennen lassen, will ich nicht alle und manche nur sehr kurz erwähnen. Da ist die Rede von Haushaltsrobotern, die Routine-Reinigungsarbeiten übernehmen, von städtischen Robotern, die den Müll einsammeln (doch die Frage, wohin damit, ist damit immer noch nicht gelöst), von selbstgesteuerten Automobilen und so banalen Wünschen wie "*no longer have to memorize phone numbers*". Mehrfach taucht das Thema "user interfaces" auf, daß also mit Hilfe der Künstlichen Intelligenz menschengerechte Benutzungsschnittstellen zu Computern geschaffen werden, daß man auch auf andere Weise als nur auf der Ebene von Fakten mit dem Computer kommunizieren kann, bis dahin, daß "*AI technologies will also be able to recognize and generate the emotional and stylistic content of such communication.*" Des weiteren wird erwähnt: "*voice recognition systems will be very good, widely implemented and will no longer be*

² siehe VDI Report 17 (1992), Abschnitt III (S. 138ff)

³ AI In the Year 2025 – Responses – comp.ai #6007

considered AI". Das haben wir schon öfter gehört, wenn etwas geklappt hat, dann ist es nicht mehr KI.

Andere Aussagen scheinen mir von größerer Tragweite zu sein: *"In the year 2025, the computer, the telephone network and television will have completely merged" [...] "the computer will become virtually (!) indistinguishable from our environment [...] And the principal key to this evolution in control over our environment is: AI!"* Es scheint bereits vom heutigen Standpunkt wahrscheinlich, daß gar nicht mehr der Computer als eine Hauptentwicklung dieses Jahrhunderts gewertet werden wird, sondern die mit dem Computer und der Kommunikationstechnologie verbundenen Netzwerke und die dadurch bedingten drastischen Einwirkungen auf gesellschaftliche Zusammenhänge. Von einem Computer, zu dem man sich zwecks Erledigung eines Auftrages begibt und wieder wegbegibt, sind wohl geringere Einflüsse auf die Lebenswelt zu erwarten, als von der medialen Vernetzung, die auch mit sich bringen wird, daß in weitem Umfang größere gesellschaftliche Gruppen weltweit miteinander in Kontakt und dadurch in neue, medial vermittelte Kommunikationsformen eintreten.

Ebenfalls bedeutsam erscheint mir die folgende Aussage: *"By the year 2025 decisionmakers of large corporations will trust their expert system over their human advisors"* – man scheint geneigt, eine Entwicklung zu akzeptieren oder mindestens für denkbar zu halten, bei der man sich mehr auf das umfangreiche Gespeicherte verläßt als auf aktuelle menschliche Perspektive und Einschätzung. Noch drastischer erscheint folgender Beitrag: *"Tapping the resources of the mind through artificial stimulation and embedded neuro-devices will be state of practice in the year 2025. Commercial offerings will include chemical, mechanical (micro machines), neural and biological approaches to enhancing decision making within the human brain – and not through external replication of the brain's processes as today"*. Möglicherweise ist diese Aussage ironisch gemeint, aber sie ist jedenfalls gegenwärtig im Denken einer Person, die sich über technologische Zukunft Gedanken macht; von einer solchen Vision ließen sich ganz andere Forschungsziele ableiten als von einer Sicht der technischen Hilfsmittel des intellektuellen Handelns, die außerhalb des menschlichen Körpers eingesetzt werden.

Jim Hendler, University of Maryland, der diese Umfrage gestartet hat, hat dann ein Statement von Freeman Dyson dazugesetzt: *"There is no illusion more dangerous than the belief that the progress of science is predictable"*. Dies paßt gut zu unserem Gedanken, daß wir keine Vorhersagen machen wollen, sondern stattdessen fiktive Szenarien diskutieren, um daran zu erörtern, was man für wünschenswert halten mag. Um zurückzukommen auf unser Thema, der möglichen Wirkungen von KI auf die Lebenswelt, sei noch eine weitere Aussage zitiert: *"The*

best AI will continue to be where the cash flow is greatest, in video-game entertainment where artificial lifeforms populate artificial reality."

Wir haben uns beispielsweise gefragt, ob neuartige Konsumgüter entstehen könnten, die die Lebenswelt drastisch verändern. Werden sie möglicherweise dazu führen, daß künstliche Welten neben einer realen Lebenswelt entstehen? Hier möchte ich nun auf unsere Überlegungen zur Verbindung von KI-Techniken mit der sog. "Virtual Reality"-Technik (auch "Artificial Reality" oder "Cyberspace") zu sprechen kommen.

Einige von Ihnen haben die Gelegenheit wahrgenommen, sich im Vorfeld des Symposiums bei der GMD erste Entwicklungsarbeiten zur technischen Nutzung von Virtual Reality demonstrieren zu lassen. Diese Technik ist ebenfalls von großer Attraktivität für die Entertainment-Industrie. Sie zielt auf eine möglichst unmittelbare Verbindung des Menschen mit seinen Aktivitäten und sensorischen Fähigkeiten mit der synthetischen Welt des Computers. Mit Hilfe der modernen Grafiktechnik soll sie eine neue qualitative Ebene des Erlebens ermöglichen; im Idealfall werden Betrachter der mit dem Rechner generierten Szenen "direkt" mit allen Sinnen in das Geschehen einbezogen, bis hin zur "Verbindung mit ihrem eigenen Intellekt" in einer künstlichen Welt. Andere Ziele, die im Zusammenhang mit der Technik der Breitbandkommunikationsnetze formuliert werden, gelten der Erschließung erweiterter Kommunikationsformen in einer virtuellen Gemeinschaft räumlich verteilter Benutzer ("Telepresence"). Bedingt durch die aktuellen Entwicklungen wurden auch schon Visionen geäußert wie: "Bonn und Berlin sind weit auseinander, man hat dann ein virtuelles Treffen in Helmstedt". Es sind jedoch auch künstliche Welten mit simulierten "Naturgesetzen" möglich, die sich radikal von unseren bisherigen Erfahrungsmöglichkeiten unterscheiden. Welche Auswirkungen dies haben kann, ist bislang kaum untersucht worden, jedoch durchaus Gegenstand von Forschungsprogrammen⁴. Berichtet werden Effekte der Seekrankheit nach virtuellen Spaziergängen, und kolportiert wird, daß manche der Beteiligten – nach der Erfahrung des Durchdringens virtueller Wände – noch drei Tage später versucht haben, durch eine reale Wand zu gehen.

Die vieldiskutierten extremen Spekulationen von Hans Moravec⁵ möchte ich hier nicht erneut aufgreifen; unsere Auseinandersetzung damit hat in folgendem kurzen Absatz gemündet, mit dessen Zitat ich meine Ausführungen hier beschließe:

⁴ s. Computer Graphics Vol. 26 No

⁵ "Mind Children", Cambridge 1988

"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 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"⁶.

Moravecs Entwürfe würden allerdings drastische Änderungen heutiger Vorstellungen vom Menschen beinhalten, und es ist allemal schwierig, sich über ihre Wünschbarkeit Gedanken zu machen.

⁶ siehe VDI Report 17 (1992), S. 167

Zukunftsvisionen der KI: Wünschbare und realistische Wirkungen auf die Arbeitswelt*

Prof. Dr. W. Coy

Coy hält die insbesondere von der Arbeitsgruppe 3 vorgenommene Trennung von fiktiven Szenarien und Vorhersagen für wenig hilfreich. Die Deklaration des Gesagten als fiktives Szenarium schwächt das Gesagte zwar ab, in der Praxis wird es aber doch als Vorhersage gelesen.

Zunächst beschreibt Coy den häufigen Metaphernwechsel, den die KI in den vergangenen Jahrzehnten durchgemacht hat. Hauptmetapher der fünfziger Jahre war das *Elektronengehirn*, eine Wortschöpfung, auf die möglicherweise die unterschweligen Reize der KI zurückzuführen sind. In den sechziger Jahren kühlen die ursprünglich an die KI geknüpften Erwartungen ab, was sich auch in einer bescheideneren Metaphernwahl niederschlägt. Es dominieren Begriffe wie Maschinensysteme oder allenfalls Maschinenfamilien. Aus den Entwicklungen der sechziger Jahre geht als Wissenschaft die Kerninformatik hervor. Fortan erklären eine heiße und eine kalte Metapher den ideologischen Hauptteil der Informatik und der KI-Landschaft. Gegen Ende der sechziger Jahre taucht - nicht ohne Einfluß des damaligen politischen Klimas - der Begriff des Werkzeugcomputers auf. Es entstehen neue Forschungsfelder wie interaktive Programmierung, Softwareergonomie, CAD, personal computing und desktop-publishing. Es verbreitet sich die Konzeption des Personalcomputers: One person, one processor. Die achtziger Jahre sind die Zeit der großflächigen Vernetzung von Computersystemen: Was organisatorisch zusammengehört, kann auch technisch zusammenwachsen. Der Computer avanciert vom Werkzeug zum Kommunikationsmedium. Gegenwärtig findet eine vollkommene Digitalisierung der bisher analogen Medien statt. Die letzte herkömmliche Schallplatte, eine immerhin hundert Jahre alte und weit verbreitete Entwicklung, wird im nächsten Jahr gepreßt, und das Verschwinden der Compactcassette ist abzusehen. Die Bedeutung des sich parallel dazu etablierenden HDTV-Fernsehens besteht weniger in der besseren Bildqualität für die Konsumenten als in der Digitalisierung der Fernsehstudios.

* Durch den Referenten freigegebene Wiedergabe seines Vortrages.

Beide konkurrierenden Metaphern, Computer als Werkzeug und Computer als Medium haben mit der Kerninformatik aber nur noch in Aspekten zu tun; als Trägerideologien sind beide Denkansätze wenig hilfreich. Wenn Prozesse um den Look- und Feelaspekt von graphical Userinterfaces und Betriebssystemen geführt werden, haben weder die Informatik noch die KI etwas Wesentliches dazu beizutragen. Insofern hängt die wissenschaftsideologische Basis der technischen wie der gesellschaftlichen Entwicklung der Informatik in nahezu peinlicher Weise hinterher. Daher ist die Frage nach dem Menschenbild in der KI bereits antiquiert. Sie beschreibt die Zukunft aus der Sicht des Jahres 1965. Anzustreben ist dagegen eine Beleuchtung der Zukunft aus der Sicht von heute. Der Computer entfaltet sein Potential, das der Golfkrieg eindrucksvoll und makaber dokumentierte, erst als Basiselement der neuen Medien, nicht aber als Basiselement der KI. So erscheint die Frage der Zeitschrift "Byte" - Is AI dead? - durchaus angemessen.

Bezogen auf die von Wachsmuth präsentierte Thematik begrüßt Coy die Beschränkung auf die technischen Aspekte und das Ausklammern der KI-inhärenten Ideologie. Allerdings taucht bei den geschilderten Szenarien und Problemfeldern die Frage auf, ob man grundsätzlich an die KI glauben muß, um die im VDI-Report gestellten Fragen zu verfolgen, oder anders: Kann die KI-Ideologie mit der Diskussion der technischen KI-Folgen ausgeblendet werden? Coy beantwortet diese Frage mit einem Zitat von Nils Bohr, wonach ein Hufeisen auch Glück bringt, wenn man nicht daran glaubt.

Bezogen auf die harte KI-These hält Coy die Feststellung, daß Maschinen partiell intelligente Leistungen vollbringen, für trivial: Natürlich können Computer schneller als Menschen bestimmte Rechenoperationen durchführen, und es ist ebenso evident, daß ein Thermostat die Heizung besser steuert als ein Hausmeister. Doch darin manifestiert sich weder Subjektivität noch Intentionalität: Thermostate denken nicht, Maschinen werden nicht zu Personen, sie tragen keine Verantwortung und können a priori nicht *ich* meinen, obschon sie es sagen können.

Als Ursache der Idee eines maschinengenerierten Geistes diagnostiziert Coy die Eigenheit des Menschen, sinnhaftes Erleben in die Objektwelt zu projizieren. Die Produktion kausaler Wirkungsketten als Sinn ist ebenso verantwortlich für die Triumphe der Wissenschaft wie für ihr zunehmendes Versagen als gesellschaftliche Leitideologie. Hier besteht eine ideengeschichtliche Linie von der Alchimie über die Astrologie und der Hegelschen Idee der Vernunft als hidden parameter der Geschichte bis zur künstlichen Intelligenz und Neuroinformatik. Überschüssige Sinnproduktion bringt den Menschen dazu, einen Eimer zu verfluchen, über den er gerade gestolpert ist, oder im PC einen Partner bzw. eine Intelligenz zu vermuten. Was die Phantasiefigur einer eigenständigen Intelligenz aufkommen läßt, ist aber nicht eine men-

schenähnliche Mimikri der Maschinen, sondern das Resultat eines permanenten, phantastischen Sinnsuchens in Zusammenhängen, in denen wir uns kognitiv bewegen. Insofern scheint die Idee der maschinellen Intelligenz ein genetischer Irrtum zu sein.

Coy kritisiert einige Details aus Wachsmuths Vortrag: Die Prognosen für die nächsten 10 Jahre haben sich zum Teil schon erfüllt. So existieren bereits automatische Sprachübersetzungssysteme. Entscheidend ist allerdings die Frage nach der Qualität. Die im Beitrag der Arbeitsgruppe 3 auftauchende Quote von 99% sprachverstehenden Robotern muß erheblich nach unten korrigiert werden. KI und Informatik verändern allerdings dramatisch die Arbeitswelt. KI und Informationsverarbeitungstechnologien wirken vor allem als Reorganisation gesellschaftlicher Arbeit. Anküpfend daran wären viele wichtige Fragen bis hin zum Sinn der abhängigen Arbeit in Industriegesellschaften zu stellen. Allerdings ist kaum eine der Fragen KI-spezifisch. In der gesamten Diskussion werden die Auswirkungen der KI überschätzt, hingegen die Folgen der medialen Vernetzung nicht angemessen berücksichtigt. Neben der medialen Vernetzung ist eine weitere Entkopplung von Produktionssystemen und Lebenswelt ebenso wahrscheinlich wie das Eintreten eines jetzt noch ausgebliebenen Rationalisierungsschubes und der Verlust einer ethisch wirksamen lokalen Einbindung. Plausibel ist ebenfalls das Entstehen einer globalen Dorfstruktur, weil durch den verstärkten Einsatz vernetzter Computermedien eine radikale Änderung der Wahrnehmungstruktur nahezu aller Menschen in nahezu allen Bereichen und Orten eintreten wird. Rechnergestützte Medien werden die Welt noch schneller verändern als es Telefon, Film usw. in diesem Jahrhundert bereits getan haben. KI ist in diesem globalen Änderungsprozeß nur ein Effekt zweiter Ordnung. Welche neuen Perspektiven ergeben sich hieraus? Es ist denkbar, daß die rationalistische Ideologie der KI oder einer ihrer geläuterten Varianten zur Modellbildung bei aufgeschlossenen Fraktionen der Biologie, Psychologie und Sozialwissenschaften beiträgt. Für die Informatik ist abzusehen, daß der Mainstream den zwar rationalistisch begründeten, aber trotzdem irrealen ideologischen Ansatz der harten KI innerhalb der derzeitigen Wissenschaftlergeneration ablegen wird, ähnlich wie die Naturwissenschaften die Astrologie überwunden haben. Im Duktus der hier diskutierten Vorhersagen bedeutet das : Weitere KI-Zentren werden nicht mehr entstehen, die bestehenden zu mehr oder weniger guten "Informatikschmieden" konvertieren oder verschwinden. KI wird in 30 Jahren nur noch mit nostalgischen Erinnerungen besetzt sein. Künstliches Leben wird nicht stattfinden.

Round-table-Gespräch: Konsequenzen für eine industrielle Technologiefolgenab- schätzung der KI

*Teilnehmer: R. Haberbeck, M.A.
Dr. R.A. Müller
Dr. G. Klett
Prof. Dr. O. Herzog*

Moderation: Dr. R. Stransfeld

Einleitung R. Stransfeld

Während die Beiträge des ersten Tages der Veranstaltung sich hauptsächlich mit theoretischen Vorstellungen aus dem Bereich der KI sowie mit diesbezüglichen Erwartungen und deren Wünschbarkeit und Verantwortbarkeit beschäftigt haben, soll dieses Round-Table Gespräch eher praktische Aspekte von KI-Entwicklung und Einsatz in den industriellen Umfeldern thematisieren. Durch diese Konfrontation mit der Praxis könnte der Gegenstand greifbarer werden, zugleich könnte diese Konfrontation auch zur Ernüchterung bezüglich hochgesteckter Erwartungen bzw. zur ernüchternden Bestätigung vorhandener Skepsis beitragen. Vielleicht werden dabei auch Lernprozesse innerhalb der Industrie sichtbar, die ja aufgrund ihrer Investitionsmöglichkeiten selbst die Wege der Entwicklung beeinflusst und daher auch die Entwicklungstendenzen kennt.

Beitrag R. Haberbeck

Der weite KI-Begriff (also derjenige, der neben "harten" symbol-logischen Ansätzen auch andere seit Mitte der fünfziger Jahre verfolgte Ansätze, z. B. aus dem Bereich der Neuroinformatik einbezieht) bezeichnet ein außerordentlich diffuses theoretisches Gebilde. Zur Formulierung theoretischer Aussagen stehen häufig nur Metaphern zur Verfügung. Obwohl diese KI eine Wissenschaft und Ingenieurdisziplin sein möchte, ist die nötige Präzision, Begriffsschärfe und axiomatische Fundierung nicht vorhanden. Deshalb läßt sich schon über die Frage, ob ein bestimmtes Produkt oder eine bestimmte Technologie in den Bereich der KI ge-

hört, unter den an der KI-Forschung und -Entwicklung Beteiligten meistens kein Konsens erzielen. Es ist also nach über dreißig Jahren noch nicht zu einem Begriffskonsens gekommen, obwohl die KI-Forschung im Zusammenhang einer zumindest einige Jahrhunderte alten Entwicklung von Wissenschaft und Wissenschaftstheorie steht.

Mit Bezug auf die Anwendung ist eine wesentliche Orientierung darüber, was im weiten Sinne KI ist, durch den militärischen Einsatz gegeben worden. Kein industrieller Bereich kann es sich leisten, über 30 Jahre hinweg Milliardenbeträge in Forschung und Entwicklung zu stecken, ohne daß zugleich profitabel vermarktbar Produkte entstehen, wenn dahinter nicht ein Glaubenskonflikt steht, wie derjenige zwischen den USA und der ehemaligen UdSSR. Alle Leitvorstellungen der KI, etwa das Ziel, in einer bestimmten Zeit eine bestimmte Anzahl von Wörtern automatisch zu erkennen, waren primär an militärischen Anwendungen orientiert. In Europa ist das aus den USA stammende KI-Paradigma naiv übernommen worden. Nach Anwendungsmöglichkeiten ist dabei selten gefragt worden, und sobald zivile Anwendungen, etwa im Bürobereich, ins Auge gefaßt wurden, ergaben sich massive Probleme. So ergaben z. B. diesbezüglich bei Nixdorf durchgeführte ergonomische Untersuchungen, daß Spracheingabe für einen Büroarbeitsplatz überhaupt nicht sinnvoll ist. Nixdorf hat daraufhin die Forschung und Entwicklung im Bereich der Spracherkennung aufgegeben.

Wenn tatsächlich die Leitbildorientierung der KI an militärischen Anwendungen dominant ist, muß man sich allerdings nicht wundern, daß es kaum Anwendungen außerhalb des militärischen Bereichs gibt. Hier spielt natürlich neben der hohen Spezifik der militärischen Aufgabenstellungen auch der militärische Verschuß eine Rolle: Interessante Entwicklungen werden zurückgehalten.

Bei der Frage nach Leitbildern und Verantwortbarkeit der KI ist in die Gegenwart und in die Zukunft zu sehen, aber die Vergangenheit sollte hier jedenfalls nicht ausgeklammert werden. Die umfangreiche und technisch betrachtet erfolgreiche Anwendung der KI im Golfkrieg sollte die KI-Gemeinde nicht von sich weisen. Ohne ihre Entwicklungsleistung und deren praktische Umsetzung hätten die USA in dieser Art und Weise keinen Krieg führen können.

Hier gibt es also ein Anwendungsbeispiel für eine sehr anspruchsvolle Entwicklung, die außerhalb der öffentlichen Kontrolle stattfand. Aber auch andere große Entwicklungsprojekte, etwa das japanische Fifth Generation Program haben nicht die erhoffte Entwicklung in Richtung auf eine allgemeine zivile Anwendbarkeit der KI gebracht (das Fifth Generation Program wird heute selbst von seinen Betreibern als Pleite betrachtet).

Die Frage ist also: Sind die Leitbilder des bloßen technischen Fortschreitens hier eigentlich ausreichend, um sinnvolle Technologien zu entwickeln, oder müssen nicht von Anfang an bei der Forschung und Entwicklung Leitbilder und Ideen existieren, die auf ein Anwendungsfeld hinzielen? Hier ist in Europa ein Umdenken erforderlich, weg vom bloßen Kopieren US-Amerikanischer Forschungsprojekte, hin zu einer an praktischen Problemen (wie sie z. B. angesichts der globalen Umweltsituation oder der sich rapide entwickelnden Kommunikationsbedürfnisse in einem zusammenwachsenden aber weiterhin vielsprachigen Europa gegeben sind) orientierten Forschung und Entwicklung auf dem gesamten Gebiet der Informationstechnologie.

Beitrag O. Herzog

Technologiefolgenabschätzung hat zwei Aspekte. Auf der einen Seite kann man fragen, welche gesellschaftlichen Folgen eine gegebene Technologie oder technologische Entwicklung in der Zukunft haben wird. Diese Frage ist sicher berechtigt und erfordert intensives Nachdenken. Auf der anderen Seite kann man jedoch auch fragen, was für Folgen eine bestimmte Technologie für Produkte und Produktionsprozesse in der Industrie hat. Die folgenden Ausführungen konzentrieren sich auf diesen zweiten Aspekt.

Es hat sich inzwischen die Auffassung durchgesetzt, daß Programmieren im wesentlichen ein Modellieren ist. Jedes Programm ist ein Beweis für einen Satz (im mathematischen Sinne), wobei die Programmierer allerdings den Satz nicht kennen, wenn das Programm nicht gut spezifiziert worden ist. Auch das ist eine Modellbildung. KI-Systeme sind Programmsysteme.

Die Debatte um die Frage "Ist die KI (oder gar die ganze Informatik) tot?" ist im Grunde genommen ein Beweis für den Erfolg der KI, denn sie wird benutzt und infolgedessen nicht mehr als KI unter diesem Label gesehen. Objektorientierte Konzepte etwa haben längst ihren Weg aus der KI-Forschung in die industrielle Anwendung als Werkzeuge zur Modellbildung angetreten. Von Objektorientiertheit ist in der industriellen Praxis heute überall die Rede, und das ist dann nur ein vornehmer Ausdruck für so etwas wie C++. Daß objektorientierte Konzepte das Modellieren sehr gut unterstützen können, ist aber in der KI-Forschung schon lange bekannt, denn hier wird die modellbasierte Entwicklung von Anwendungen schon seit mehreren Jahrzehnten betrieben. Diejenigen Konzepte der KI, die theoretisch wirklich beherrscht werden, dazu gehören etwa bestimmte Wissensrepräsentationssprachen oder Inferenzmaschinen, finden dann sehr schnell ihren Weg in die industrielle Anwendung, und kein Mensch merkt mehr von außen, daß in einem solchen System KI steckt.

Zu den wichtigen Domänen der industriellen Anwendung von KI Technologie gehören heute neben der Softwareentwicklung vor allem die Bereiche der Repräsentation von formalisierbarem Wissen (Expertensysteme lassen sich heute z. B. im Bereich der Ökologie mit großem Nutzen einsetzen) und der Benutzerschnittstellen (im Unterschied zu der von Haberbeck referierten Entwicklung bei Nixdorf hat IBM z. B. seine Aktivitäten im Bereich der Spracherkennung nicht aufgegeben). Natürlich wird z. B. im Bereich der Spracherkennung nicht nur mit KI, sondern auch mit anderen, etwa statistischen Methoden gearbeitet. Sobald allerdings semantische Aspekte einbezogen werden, kommt man direkt wieder zu KI-Konzepten. Allerdings wird kein Mensch, der etwa mit einem solchen System Texte erfaßt, dieses als ein KI-System wahrnehmen. Es wird schlicht als ein Werkzeug betrachtet.

Es wird also auch in der industriellen Technologie, in Produkten, KI-Technologie geben, und zwar auf der einen Seite, um die Benutzerschnittstellen wirklich benutzerfreundlich zu machen, und auf der anderen Seite im Kern von Systemen, um Wissen zu repräsentieren. Das ist, was man heute kann, und deshalb sagen viele, die KI sei tot, weil man nicht mehr sieht, was da KI ist. Dabei wird allerdings ein wesentlicher Punkt übersehen: Das, was heute in Softwareprodukte eingeht, sind nur ganz kleine Teilgebiete der Forschung und Entwicklung auf dem Gebiet der KI. Viele KI-Forscher forschen ja bereits auf ganz neuen Gebieten, und diese Ergebnisse gehen dann vielleicht in 20 Jahren in Produkte ein. Als Kernaussage ergibt sich, daß es einen ganz konstanten Fluß von Technologie gibt, der aus der KI (und aus der Informatik allgemein) in Anwendungen übergeht, und dieser Fluß wird sicherlich nicht aufhören, und deshalb wird das, was wir mit dem Schlagwort KI bezeichnen, noch lange leben.

Beitrag G. Klett

Die Erwartungen, die sich 1985 mit der Einführung der KI-Technologie bei der BASF verbanden, entsprechen nicht dem Bild, welches in der ersten These des Beitrags von Frau Krämer gezeichnet wird (KI sei eher eine Ideologie denn eine ingenieurwissenschaftliche Bereicherung). Damals schien am Markt eine Methode aufzukommen, die es gestatten sollte, den Fachexperten Werkzeuge an die Hand zu geben, um ihre Modelle, die sie im Hinterkopf (oder auch nur im Bauch) haben, leichter für sich selber artikulieren zu können, als dies bisher mit Hilfe eines Systemanalytikers von der Fachinformatik möglich war. Damit verband sich die Hoffnung, daß die Fachabteilungen (Elektrotechnik, Prozessleittechnik, Verfahrenschmie, usf.) ihre anstehenden Probleme selber formulieren und im Sinne einer evolutionären Programmierung Lösungsmodelle schnell ausprobieren können würden.

In dieser Situation wurde anfänglich fast jedes Tool, das am Markt war, untersucht. Dabei wurden unter anderem auch leidvolle Erfahrungen mit der dazu erforderlichen Spezialhardware gemacht. Es wurden ungefähr 20 Projekte aufgelegt (von Designsystemen für Katalysatoren bis zu Expertensystemen, die an Prozeßleitständen als *operator advisory* eingesetzt werden sollten). Es war sicherlich ausreichender Ehrgeiz vorhanden, und es waren auch Leute da, die sich mit der Idee identifizieren konnten, also die von Wachsmuth angesprochenen Konsumpioniere.

Heute läßt sich jedoch das Resümee ziehen, daß KI bei der BASF als eigenständige Systemlösung nicht existiert, sondern höchstens als Teil einer Gesamtlösung, in die KI-Komponenten integriert werden. Des weiteren werden KI- (oder wissensbasierte) Systeme nur noch mit einer Art Beipackzettel ausgeliefert, der Nebenwirkungen und Kontraindikationen angibt. Insbesondere werden alle diesbezüglichen Projekte vorab daraufhin geprüft, ob Spezialhardware (Rechnersysteme, Spezialprozessoren, *add-ons*, Karten) gebraucht werden. Wenn das der Fall ist, wird ein Projekt nicht durchgeführt. In der gleichen Weise wird softwareseitig Wert auf die Verwendung von standardisierten Oberflächen, Standard-Datenbanken, Standard-Compilern, usf. gelegt. Zugleich wird aufgrund entsprechender negativer Erfahrungen geprüft, ob es gravierende Unterschiede zwischen der Laufzeitumgebung und der Entwicklungsumgebung gibt. Nur wenn keine dieser Kontraindikationen gegeben ist und wenn sich das Problem nicht mit vertretbarem Aufwand auch prozedural lösen läßt, werden wissensbasierte Systeme eingesetzt.

Ein positiver Aspekt hat bei der Beschäftigung mit wissensbasierten Systemen vor allem darin bestanden, daß viele Fachexperten durch den Zwang, ihr Problem zu artikulieren, ihr eigenes Wissen einer kritischen Reflektion unterziehen mußten. Der Einsatz wissensbasierter Methoden und Werkzeuge erfordert eine formal explizite, konstruktive Darstellung des Expertenwissens, das normalerweise nur in Form von Versatzstücken in den Köpfen der Fachexperten existiert. Man kann dann auch sofort in einem Prototyp sehen, was man da eigentlich artikuliert hat. Das führt dann in der Regel zu einem Überdenken und Abändern der ursprünglichen Konstruktion. Auf diese Weise haben viele Fachexperten, auch wenn sie schließlich kein Expertensystem einsetzen, ihr Wissen teilweise sehr stark verändert und verfeinert. Ein Beispiel für diesen Vorgang hat sich kürzlich ergeben, als mit einem Fuzzy-Regler für eine Destillationskolonne experimentiert wurde: Der Mitarbeiter, der die Kolonne früher von Hand gefahren hatte, macht das heute ganz anders, weil er bei der Erprobung des Fuzzy-Reglers auf ganz neue Möglichkeiten und Wege gestoßen ist.

Zusammenfassend läßt sich jedoch sagen, daß bei der BASF KI-Methoden nur auf Sparflamme weiterverfolgt werden, daß der große Durchbruch nicht stattgefunden hat. Dies liegt zum Teil daran, daß die Erwartungen, die an solche Systeme geknüpft wurden, zu hochgesteckt waren, und zum Teil an der oft mangelhaften Unterstützung der Hard- und Softwarehersteller auf diesem Gebiet. Bei den Systemen, die laufen und im ständigen Einsatz akzeptiert werden, handelt es sich durchweg um kleinere Systeme, die auf einer breiten Rechnerbasis (überwiegend auf PCs) laufen. Diese Systeme sind in größere Systeme integriert und stellen keine eigenständigen Lösungen dar. Dieser Weg wird in Zukunft weiter beschritten werden. Im Augenblick gibt es darüberhinaus einige interessante und vielversprechende Entwicklungen im Bereich der Fuzzy-Technologie (die man als eine Renaissance des Analogrechners betrachten kann). Für die nähere Zukunft ist jedoch auf der Basis des derzeitigen Standes der Technik und der Erfahrungen mit existierenden Systemen nicht davon auszugehen, daß die KI-Bereiche wie etwa die *management information systems* revolutioniert.

Beitrag R. A. Müller

In dem folgenden Statement geht es in erster Linie um KI im Unternehmen, nur soweit sie in Herstellungsprozessen im weiteren Sinne (d. h. in Verfahren der Beschaffung, der Produktion, des Vertriebs, der Planung, der Administration, usf.) eine Rolle spielt. KI in Produkten (etwa in Fahrzeugen oder Flugzeugen) soll hier ausgeklammert bleiben. Die folgenden drei Thesen beziehen sich auf die Relevanz der KI im Unternehmen, Leitbilder der KI in diesem Zusammenhang und Technologiefolgenabschätzung.

- Relevanz: Die vielfach zu hörenden Warnungen vor der KI, bei denen häufig gleich das Abendland in Gefahr gesehen wird, sind mit Bezug auf den hier infragestehenden Bereich zu relativieren. Tatsächlich wird kaum irgendwo ein KI-Projekt erfolgreich abgeschlossen bzw. ein KI-System erfolgreich eingesetzt, wenn man als Maßstab des Erfolges nicht die erfolgreiche Aquisition eines Projektes oder seine schließliche Abnahme ansetzt, sondern den Nachweis einer Qualitätsverbesserung oder einer Kostensenkung durch den Einsatz eines KI-Systems ansetzt. Setzt man dieses Kriterium an, dürften 80 Prozent aller installierten Software durch den Rost fallen. Ohnehin ist der KI-Anteil an der gesamten heute in Unternehmen installierten Software marginal. Eine Bedeutung hat KI hier höchstens in der Hinsicht, daß einzelne Entwicklungen aus der KI in bereits etablierte Softwareprodukte eindiffundieren. KI erscheint also immer eingebettet in bestehende Anwendungszusammenhänge, von einer größeren Relevanz, geschweige denn einer Gefahr, die von dieser Technologie derzeit ausgeht, kann daher kaum die Rede sein.

- Leitbilder: Das Leitbild der klassischen KI, nämlich KI als Nachbildung menschlicher Fähigkeiten, paßt nicht in das Managementkonzept moderner Unternehmen. Der "künstliche Experte" ist eine Metapher, die in der Realität nicht in Erscheinung tritt, nicht nur unerwünscht ist, sondern wahrscheinlich sogar schädlich wäre; entsprechende Projekte scheitern dementsprechend regelmäßig. Hingegen besteht im Unternehmen ein Bedarf an Unterstützungssystemen, die an Arbeitsprozesse optimal und flexibel anpaßbar sind und die zu einer besseren Vernetzung nicht von Daten, sondern von menschlichen Kompetenzen führen. KI sollte der besseren Verzahnung heterogener Arbeitsprozesse dienen. Das ist eine große Herausforderung und deckt sich genau mit der Sicht des Computers als Medium, wie sie von Coy skizziert wurde. Das Leitbild des KI-Systems als "Problemlöser" ist in diesem Zusammenhang von eher untergeordneter Bedeutung. Hauptsächlich geht es um die sinn-erhaltende Umformung von Texten, Grafiken, usf. also in einem weiten Sinne um "Übersetzungen". Es muß nicht gleich um so gewaltige Dinge wie Problemlösungen gehen. Die Übersetzung etwa einer Stückliste von einer textuellen in eine graphische oder gemischte Form oder in eine multimediale Anwendung ist bereits eine genügend hohe Herausforderung, ohne daß dabei im Sinne der klassischen KI Probleme gelöst würden. Betont wird hier die kommunikative Funktion des KI-Systems.

Es gibt eine Alternative zur vorherrschenden Strategie der maschinenorientierten und technologiegetriebenen Entwicklung, bei der eine neue Technologie aus der Grundlagenforschung, etwa von den Universitäten, kommt und die Industrie dann aufgefordert ist, auf dieser Basis neue Produkte zu entwickeln und an den Markt zu bringen. Die Alternative ist eine arbeitsprozessorientierte und benutzergetriebene Entwicklung. Ziel ist es, den individuellen Arbeitsprozeß mit seinen Schnittstellen zu anderen Prozessen zu effektivieren und die Systemgestaltung bereits in der Forschung von den Benutzern triggern zu lassen. Überspitzt gesagt stehen hier zwei Sichtweisen einander gegenüber: Zum einen die Sicht des Unternehmens als einer riesigen Maschine, in der der einzelne Mensch die Rolle eines Maschinenbedieners hat, zum anderen die Sicht des Unternehmens als einem sozialen System, bestehend aus Gruppen von Menschen, die Werkzeuge benutzen. Würde man solche meist nur implizit gegebenen Leitbilder stärker bewußt machen, würden sich auch andere KI-Paradigmen herausdifferenzieren. Dem technozentrischen Leitbild der Forschung, welches darauf hinausläuft, die Maschine immer intelligenter zu machen, könnte man ein anthropozentrisches Leitbild entgegenstellen, bei dem es nicht darauf ankommt, die Maschine intelligenter zu machen, sondern die Organisation lebensfähiger. Diese beiden Leitbilder stehen sich allerdings nicht antipodisch gegenüber, sondern bilden eher ein Vexierbild.

- **Technologiefolgenabschätzung:** Es ist stark zu bezweifeln, daß es KI-spezifische Technologiefolgen gibt. KI-Systeme sind, wie bereits in verschiedenen Beiträgen gezeigt wurde, hinsichtlich ihrer Wirkung im Unternehmen von konventionellen Informationstechniken nicht unterscheidbar. Das, worin sich die KI von der konventionellen Informationstechnologie unterscheidet, sind z. B. bestimmte Programmierstile, die in bestimmten Anwendungsklassen effizientere Programmierung erlauben (z. B. objektorientiertes Programmieren) und bestimmte Programmarchitekturen, die in bestimmten Bereichen mächtigere Softwarefunktionalitäten, z. B. Inferenz, zur Verfügung stellen.

- **Das Design von Unterstützungssystemen** sollte von der angestrebten Wirkung dieser Systeme im Arbeitsprozeß ausgehen. Es handelt sich dann weniger um ein KI-Problem als vielmehr um ein organisatorisches bzw. arbeitswissenschaftliches Problem und natürlich auch um ein Problem derer, die arbeiten. Die Anforderungen würden also primär von den Benutzern bzw. stellvertretend von deren Management bestimmt und nicht von der KI-Forschung. Bei einem solchen Modell ist die Betrachtung der Technologiefolgen methodisch bereits im Technikgenese-prozeß verankert, und dann ist *cum grano salis* eine zusätzliche, externe Technologiefolgen-Abschätzung überhaupt nicht notwendig.

Die im Anschluß daran durchgeführte Diskussion ist im Diskurs - Protokoll zur Technikfolgenabschätzung der Informationstechnik II - 3 veröffentlicht (Hrsg.: VDI/VDE-Technologiezentrum Informationstechnik GmbH, Dezember 1993, S. 31-34).

Anhang

Mitglieder des VDI-Ausschusses "Künstliche Intelligenz" und Gastreferenten

Prof. Dr. Armin B. Cremers
(Obmann)

Rheinische Friedrich-Wilhelms-Universität Bonn
Institut für Informatik III
Römerstr. 164
53117 Bonn

Prof. Dr. Gerhard Barth
(Gastreferent)

Daimler-Benz AG
Forschung und Technik
Abteilung F 3
Wilhelm-Runge-Str. 11
89081 Ulm
(vormals Deutsches Forschungszentrum
für KI, Kaiserslautern)

Dr. Barbara Becker

GMD Birlinghoven
Institut für angewandte Informationstechnik
Forschungsgruppe Expertensysteme
Postfach 12 40
53730 Sankt Augustin

Prof. Dr. Rafael Capurro

Fachhochschule für Druck
Feuerbacher Heide 38-42
70192 Stuttgart

Prof. Dr. Wolfgang Coy
(Gastreferent)

Universität Bremen
Fachbereich Mathematik/Informatik
Postfach 33 04 40
28334 Bremen

Prof. Dr. Meinolf Dierkes
(Gastreferent)

Wissenschaftszentrum Berlin
für Sozialforschung
Reichpietschufer 50
10785 Berlin

Prof. Dr.-Ing. Rolf Eckmiller

Rheinische Friedrich-Wilhelms-Universität Bonn
Institut für Informatik VI
Römerstr. 164
53117 Bonn
(vormals Universität Düsseldorf)

Prof. Dr. Günther Görz

Universität Erlangen-Nürnberg
IMMD 8 - KI
Am Weichselgarten 9
91058 Erlangen

Rolf Haberbeck M.A.

SNI
Gustav-Meyer-Allee 1
13355 Berlin

Prof. Dr. Otthein Herzog
(Gastreferent)

Universität Bremen
Fachbereich Informatik
Arbeitsgemeinschaft KI
Postfach 33 04 40
28334 Bremen
(vormals IBM Deutschland GmbH, Stuttgart)

Prof. Dr. Andreas Kemmerling

Ludwig-Maximilians-Universität München
Institut für Statistik
Geschwister-Scholl-Platz 1
80539 München

Dr. Gerhard Klett
(Gastreferent)

BASF AG
Abteilung Zentrale Informatik-Technologie
Kaiser-Wilhelm-Str. 52
67059 Ludwigshafen

Prof. Dr. Sybille Krämer	Freie Universität Berlin Institut für Philosophie Habelschwerdter Allee 30 14195 Berlin
Dr. Anton Kremeier	Schinkelstr. 30 c 40699 Erkrath
Prof. Dr.-Ing. Jan Lunze (Gastreferent)	Technische Universität Hamburg-Harburg Arbeitsbereich Regelungstechnik Eißendorfer Str. 40 21073 Hamburg
Dr. Rolf A. Müller	Daimler-Benz AG Ressort Forschung und Technik Forschungsinstitut Berlin Alt-Moabit 91 b 10557 Berlin
Dr. Horst Röpke	Schering AG Postfach 65 03 11 13342 Berlin
Prof. Dr. Andreas Schlachetzki	Technische Universität Braunschweig Institut für Halbleitertechnik Postfach 33 29 38106 Braunschweig
Dr. Paul Schreiber	Bundesanstalt für Arbeitsschutz Vogelpothsweg 50-52 44149 Dortmund
Dr. Jürgen Seetzen	VDI/VDE-Technologiezentrum Informationstechnik GmbH Rheinstr. 10 b 14513 Teltow

Prof. Dr. Jörg Siekmann
(Gastreferent) Deutsches Forschungszentrum für
Künstliche Intelligenz
Stuhlsatzenhausweg 3
66123 Saarbrücken

Dr. Reinhard Stransfeld
(Gastreferent) VDI/VDE-Technologiezentrum
Informationstechnik GmbH
Rheinstr. 10 b
14513 Teltow

Prof. Dr. Ipke Wachsmuth Universität Bielefeld
Technische Fakultät
Wissensbasierte Systeme
Postfach 10 01 31
33501 Bielefeld

Prof. Dr. Karl F. Wender Universität Trier
Fachbereich 1
Psychologie
Postfach 38 25
54286 Trier

Wissenschaftliche Mitarbeiter

Matthias Butt Bundesallee 90
12161 Berlin

Dipl.-Soz. Michael Wilker Am Osterberg 14
49152 Bad Essen

Gernot Grube Bautzener Str. 17
10289 Berlin

Aktuelle Veröffentlichungen aus der VDI-Hauptgruppe zum Thema "Mensch und Technik"

Die folgenden Veröffentlichungen sind über die VDI-Hauptgruppe bzw. über den VDI-Verlag Vertriebsleitung, Postfach 10 10 54, 40001 Düsseldorf zu erhalten.

Kurt A. Detzer: Von den zehn Geboten zu Verhaltenskodizes für Manager und Ingenieure. Was sagen uns ethische Leitbilder, 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)

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

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

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

Technikbewertung - Begriffe und Grundlagen. Erläuterungen und Hinweise zur VDI-Richtlinie 3780 (VDI-Report 15, Dezember 1991) (Hrsg.: VDI-Hauptgruppe)

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. Stuttgart (Schäffer-Poeschel-Verlag) 1994

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, Ingenieur-Gehälter 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“

DM 20,00

VDI-Report 20

„Hochschulbildung und Ingenieurberuf“

DM 20,00

VDI-Report 21

„Künstliche Intelligenz“ – Tagungsbericht

DM 18,00

VDI-Report 22

„Technik und Kunst – Berührungspunkte
und Abgrenzungen in historischer Perspektive“

DM 30,00

