

Einleitung

Günther Görz und Ipke Wachsmuth

„Künstliche Intelligenz“ (KI) ist eine wissenschaftliche Disziplin, die das Ziel verfolgt, menschliche Wahrnehmungs- und Verstandesleistungen zu operationalisieren und durch Artefakte, kunstvoll gestaltete technische – insbesondere informationsverarbeitende – Systeme verfügbar zu machen². Unter den zahlreichen Definitionen der Disziplin, die von ihren Fachvertretern angegeben wurden, sei beispielhaft die von [Winston 1992] formulierte genannt, die die Bestimmung des Gegenstands der KI in folgender Weise präzisiert:

„Künstliche Intelligenz ist die Untersuchung von Berechnungsverfahren, die es ermöglichen, wahrzunehmen, zu schlußfolgern und zu handeln.“

Diese Aufgabenstellung impliziert den grundsätzlich interdisziplinären Charakter der KI: Obwohl durch ihre Genese zumeist in der Informatik als Teilgebiet verankert und damit ihre ingenieurwissenschaftliche Komponente betonend, ist KI-Forschung gleichwohl nur in enger Zusammenarbeit mit Philosophie, Psychologie, Linguistik und den Neurowissenschaften möglich, die für ihre kognitionswissenschaftliche Komponente grundlegend sind. So unterscheidet sich die KI von der klassischen Informatik wegen der Betonung von Wahrnehmung, Schlußfolgern und Handeln, und sie unterscheidet sich von der Psychologie wegen der Betonung des Aspekts der Berechnung. Ein Kernpunkt dabei ist die These, daß das „Räsonnieren“ konstitutiv für höhere Intelligenzfunktionen ist. Schlußfolgerndes Denken – in einem sehr allgemeinen Sinn – involviert interne Prozesse, die es einem Individuum ermöglichen, darüber nachzudenken, was die beste Weise zu handeln ist, bevor tatsächlich gehandelt wird. Entscheidend dabei ist der Rückgriff auf Wissen über die Welt und über alternative Möglichkeiten des Handelns in der Welt. Die Bezeichnung „Künstliche Intelligenz“ ist historisch zu verstehen: Zunächst im Englischen als „Artificial Intelligence“ geprägt, ist sie als wörtliche Übersetzung nicht sinngemäß und gibt Anlaß zu dem Mißverständnis, sie würde eine Definition von „Intelligenz“ liefern oder hätte gar einen operationalisierbaren Intelligenzbegriff insgesamt zu entwickeln. Da die KI eine junge Disziplin ist, zeichnet sich ihre Grundlagendiskussion zudem durch eine metaphernreiche und aufgrund ihres Gegenstands auch stark anthropomorphe Sprache aus.

Dennoch bleibt uns eine grundsätzlichere Auseinandersetzung mit dem Begriff der *Intelligenz* nicht erspart. „Intelligenz ist die allgemeine Fähigkeit eines Individuums, sein Denken bewußt auf neue Forderungen einzustellen; sie ist allgemeine geistige Anpassungsfähigkeit

²Mit Kant wollen wir unter „Verstand“ das Vermögen der Regeln verstehen – im Unterschied zur Vernunft als Vermögen der Prinzipien.

an neue Aufgaben und Bedingungen des Lebens.“ Diese noch recht unpräzise Bestimmung durch den Psychologen William Stern aus dem Jahre 1912 hat eine Vielzahl von Versuchen nach sich gezogen, eine zusammenhängende Intelligenztheorie zu erstellen, deren keiner dem komplexen Sachverhalt auch nur annähernd gerecht werden konnte [Irrgang, Klawitter 1990]. Heute besteht weitgehend Konsens darüber, daß Intelligenz zu verstehen ist als Erkenntnisvermögen, als Urteilsfähigkeit, als das Erfassen von Möglichkeiten, aber auch als das Vermögen, Zusammenhänge zu begreifen und Einsichten zu gewinnen.

Sicherlich wird Intelligenz in besonderer Weise deutlich bei der Fähigkeit, Probleme zu lösen. Die Art, die Effizienz und die Geschwindigkeit, mit der sich der Mensch bei der Problemlösung an die Umwelt anpaßt (Akkommodation) oder die Umwelt an sich angleicht (Assimilation), ist ein wichtiges Merkmal von Intelligenz. Dabei äußert sich Intelligenz durchaus nicht nur in abstrakten gedanklichen Leistungen wie logischem Denken, Rechnen oder Gedächtnis und insbesondere in der Fähigkeit zur Reflexion, sondern wird ebenso offenkundig beim Umgang mit Wörtern und Sprachregeln oder beim Erkennen von Gegenständen und Situationsverläufen. Neben der konvergenten Fähigkeit, eine Vielzahl von Informationen zu kombinieren, um dadurch Lösungen zu finden, spielt bei der Problemlösung aber auch die *Kreativität* eine wichtige Rolle, insbesondere auch das Vermögen, außerhalb der aktuellen Informationen liegende Lösungsmöglichkeiten einzubeziehen. Andererseits ist gerade die Fähigkeit zur Begrenzung der Suche nach Lösungen bei hartnäckigen Problemen eine typische Leistung der Intelligenz.

Und all dies, so müssen wir an dieser Stelle fragen, soll Gegenstand einer künstlichen Intelligenz sein? Kurz gesagt: nein, denn schon wenn wir Intelligenz beurteilen oder gar messen wollen, bedarf es einer Operationalisierung, wodurch wir einen Übergang vom personalen Handeln zum schematischen, nicht-personalen Operieren vollziehen. Das, was operationalisierbar ist, läßt sich grundsätzlich auch mit formalen Systemen darstellen und damit auf einem Computer berechnen. Vieles aber, was das menschliche Denken kennzeichnet und was wir mit intentionalen Termini wie Kreativität oder Bewußtsein benennen, entzieht sich weitgehend einer Operationalisierung. Dies wird jedoch angezweifelt von Vertretern der sog. „starken KI-These“, die besagt, daß Bewußtseinsprozesse *nichts anderes* als Berechnungsprozesse sind, die also Intelligenz und Kognition auf bloße Informationsverarbeitung reduziert. Ein solcher Nachweis konnte aber bisher nicht erbracht werden – die Behauptung, es sei *im Prinzip* der Fall, kann den Nachweis nicht ersetzen. Hingegen wird kaum bestritten, daß Intelligenz *auch* Informationsverarbeitung ist – dies entspricht der „schwachen KI-These“.

So, wie wir Intelligenz erst im sozialen Handlungszusammenhang zuschreiben, ja sie sich eigentlich erst darin konstituiert, können wir dann allerdings auch davon sprechen, daß es – in einem eingeschränkten Sinn – Intelligenz in der Mensch-Maschine-Interaktion, in der Wechselwirkung gibt, als „Intelligenz für uns“. Es besteht gar keine Notwendigkeit, einem technischen System, das uns als Medium bei Problemlösungen unterstützt, Intelligenz *per se* zuzuschreiben – die Intelligenz manifestiert sich in der Interaktion.

Eine generative Theorie der Intelligenz

Zentrales wissenschaftliches Ziel der KI ist es, zu bestimmen, welche Annahmen über die Repräsentation und Verarbeitung von Wissen und den Aufbau von Systemen die verschiedenen Aspekte der Intelligenz erklären können [Winston 1992]. Der vorherrschende Gedanke in den verschiedenen theoretischen Fassungen des Intelligenzbegriffs in der KI ist, daß Intelligenz aus der Interaktion vieler einfacher Prozesse „im Konzert“ emergiert und daß Prozeßmodelle intelligenten Verhaltens mit Hilfe des Computers im Detail untersucht werden können. Die Organisation des Zusammenwirkens verschiedener Softwarekomponenten, die bestimmte Teilaufgaben versehen, die Systemarchitektur, ist also ein wichtiges Thema. Sie legt die Voraussetzungen für die Entstehung von Synergieeffekten: das Zusammenspiel vieler – manchmal relativ einfacher – Komponenten kann komplexes Verhalten bewirken.

„Künstliche Intelligenz“ ist ein synthetischer Begriff, der – vermöge seines suggestiven Potentials – viele Mißverständnisse und falsche Erwartungen verursacht hat. Sein Ursprung läßt sich auf das Jahr 1956 zurückverfolgen, ein Jahr, das in vielerlei Hinsicht bedeutsam war. Zum Beispiel erschien in diesem Jahr das Buch „Automata Studies“ mit einer Reihe heute berühmter Artikel im Gebiet der Kybernetik [Shannon, McCarthy 1956]. Ebenfalls in diesem Jahr erhielten Bardeen, Shockley und Brattain den Nobelpreis für die Erfindung des Transistors. Noam Chomsky war im Begriff, seinen berühmten Artikel über syntaktische Strukturen zu veröffentlichen, der den Weg für eine theoretische Betrachtung der Sprache eröffnete [Chomsky 1957].

Die Bezeichnung „Artificial Intelligence“ wurde von John McCarthy als Thema einer Konferenz geprägt, die im Sommer 1956 am Dartmouth College stattfand und an der eine Reihe renommierter Wissenschaftler teilnahmen (u.a. Marvin Minsky, Nathaniel Rochester, Claude Shannon, Allan Newell, Herbert Simon). Dieses Treffen wird allgemein als Gründungsereignis der Künstlichen Intelligenz gewertet. Im Förderungsantrag an die Rockefeller-Stiftung wurde ausgeführt (s. [McCorduck 1979], S. 93):

„Wir schlagen eine zweimonatige Untersuchung der Künstlichen Intelligenz durch zehn Personen vor, die während des Sommers 1956 am Dartmouth College in Hannover, New Hampshire, durchgeführt werden soll. Die Untersuchung soll auf Grund der Annahme vorgehen, daß jeder Aspekt des Lernens oder jeder anderen Eigenschaft der Intelligenz im Prinzip so genau beschrieben werden kann, daß er mit einer Maschine simuliert werden kann.“

Es geht, so McCarthy später, um die „Untersuchung der Struktur der Information und der Struktur von Problemlösungsprozessen, unabhängig von Anwendungen und unabhängig von ihrer Realisierung“. Und Newell: „Eine wesentliche Bedingung für intelligentes Handeln hinreichender Allgemeinheit ist die Fähigkeit zur Erzeugung und Manipulation von Symbolstrukturen. Zur Realisierung symbolischer Strukturen sind sowohl die Instanz eines diskreten kombinatorischen Systems (lexikalische und syntaktische Aspekte), als auch die Zugriffsmöglichkeiten zu beliebigen zugeordneten Daten und Prozessen (Aspekte der Bezeichnung, Referenz und Bedeutung) erforderlich.“

Als Instrument der Forschung sollte der Universalrechner dienen, wie Minsky begründete:

„... weil Theorien von mentalen Prozessen zu komplex geworden waren und sich zu schnell entwickelt hatten, als daß sie durch gewöhnliche Maschinerie realisiert werden konnten. Einige der Prozesse, die wir untersuchen wollen, nehmen substantielle Änderungen in ihrer eigenen Organisation vor. Die Flexibilität von Computerprogrammen erlaubt Experimente, die nahezu unmöglich in 'analogen mechanischen Vorrichtungen' wären“.

Im September 1956 fand am Massachusetts Institute eine zweite wichtige Konferenz statt, das „Symposium on Information Theory“. So, wie die KI ihren Ursprung auf die Dartmouth Conference zurückführt, kann dieses Symposium als Grundsteinlegung der Kognitionswissenschaft gelten (vgl. [Gardner 1985], Kap. 3). Unter den Teilnehmern beider Konferenzen waren Allen Newell und Herbert Simon. Zusammen mit John Shaw hatten sie gerade die Arbeiten an ihrem „Logic Theorist“ abgeschlossen, einem Programm, das mathematische Sätze aus Whiteheads und Russells „Principia Mathematica“ beweisen konnte. Dieses Programm verkörperte schon, was später der Informationsverarbeitungs-Ansatz des Modellierens genannt wurde. Der Grundgedanke dieses Ansatzes ist, daß Theorien des bewußten menschlichen Handelns auf der Basis von Informationsverarbeitungs-Systemen formuliert werden, also Systemen, die aus Speichern, Prozessoren und Steuerstrukturen bestehen und auf Datenstrukturen arbeiten. Seine zentrale Annahme besteht darin, daß im Hinblick auf intelligentes Verhalten der Mensch als ein solches System verstanden werden kann.

Ein Charakteristikum für das methodische Vorgehen der KI als akademische Disziplin ist, menschliche Intelligenz dadurch zu verstehen und zu erklären, daß es gelingt, Effekte der Intelligenz – nämlich intelligentes Verhalten – zu produzieren. Fortschritte werden aufgrund lauffähiger Systeme angestrebt: Synthese vor Analyse. Es ist nicht das Ziel, intelligente Systeme zu konstruieren, *nachdem* ein Verständnis menschlicher Intelligenz erlangt wurde, sondern menschliche Intelligenz *durch* die Konstruktion solcher Systeme verstehen zu lernen.

Die Entwicklung der KI

In der Folge der genannten Tagungen wurden an verschiedenen universitären und außer-universitären Einrichtungen einschlägige Forschungsprojekte ins Leben gerufen. Die Prognosen waren zunächst optimistisch, ja geradezu enthusiastisch: Die Künstliche Intelligenz sollte wesentliche Probleme der Psychologie, Linguistik, Mathematik, Ingenieurwissenschaften und des Managements lösen. Fehlschläge blieben nicht aus: So erwies sich das Projekt der automatischen Sprachübersetzung, dessen Lösung man in greifbarer Nähe sah, als enorm unterschätzte Aufgabe. Erst in den letzten Jahren wurde es – allerdings mit größerer Bescheidenheit – wieder in Angriff genommen.

In den zahlreichen Versuchen, die Aufgaben der Disziplin zu formulieren, werden zwei Aspekte deutlich, in denen kognitionswissenschaftliche und ingenieurwissenschaftliche Zielsetzungen zum Tragen kommen:

- *Kognitive Modellierung*, d.h. Simulation kognitiver Prozesse durch Informationsverarbeitungsmodelle;

- Konstruktion „intelligenter“ Systeme, die bestimmte menschliche Wahrnehmungs- und Verstandesleistungen maschinell verfügbar machen.

In der Entwicklung der Künstlichen Intelligenz kann man vier Phasen unterscheiden: Die Gründungsphase Ende der fünfziger und Anfang der sechziger Jahre, gekennzeichnet durch erste Ansätze zur symbolischen, nicht-numerischen Informationsverarbeitung, beschäftigte sich mit der Lösung einfacher Puzzles, dem Beweisen von Sätzen der Logik und Geometrie, symbolischen mathematischen Operationen, wie unbestimmter Integration, und Spielen wie Dame und Schach. Das Gewicht lag darauf, die grundsätzliche technische Machbarkeit zu zeigen. In dieser ersten Phase – oft durch die Bezeichnung „Power-Based Approach“ charakterisiert, erwartete man sehr viel von allgemeinen Problemlösungsverfahren, deren begrenzte Tragweite allerdings bald erkennbar wurde.

Die zweite Entwicklungsphase der KI ist gekennzeichnet durch die Einrichtung von Forschungsgruppen an führenden amerikanischen Universitäten, die begannen, zentrale Fragestellungen der Künstlichen Intelligenz systematisch zu bearbeiten, z.B. Sprachverarbeitung, automatisches Problemlösen und visuelle Szenenanalyse. In dieser Phase begann die massive Förderung durch die „Advanced Research Projects Agency“ (ARPA) des amerikanischen Verteidigungsministeriums.

In den siebziger Jahren begann eine dritte Phase in der Entwicklung der KI, in der u.a. der Entwurf integrierter Robotersysteme und „expertenhaft problemlösender Systeme“ im Mittelpunkt stand. Letztere machten Gebrauch von umfangreichen codierten Wissensbeständen über bestimmte Gebiete, zunächst in Anwendungen wie symbolische Integration oder Massenspektroskopie. Im Gegensatz zum „Power-Based Approach“ trat die Verwendung formalisierten Problemlösungswissens und spezieller Verarbeitungstechniken in den Vordergrund, was durch die Bezeichnung „Knowledge-Based Approach“ charakterisiert wird. Durch diese Schwerpunktsetzung wurden große Fortschritte bei Techniken der Wissensrepräsentation und in der Systemarchitektur, besonders im Hinblick auf Kontrollmechanismen, erzielt. Im weiteren Verlauf wurde erhebliches Gewicht auf komplexe Anwendungen gelegt: Erkennung kontinuierlich gesprochener Sprache, Analyse und Synthese in der Chemie, medizinische Diagnostik und Therapie, Prospektion in der Mineralogie, Konfiguration und Fehleranalyse technischer Systeme. Zu dieser Zeit waren auch in Europa, vor allem in Großbritannien und Deutschland, KI-Forschungsgruppen an verschiedenen Universitäten entstanden und Förderprogramme installiert.

Der Eintritt in die vierte Entwicklungsphase der KI erfolgte um 1980, die vor allem durch eine umfassende Mathematisierung des Gebiets, eine Präzisierung des Konzepts der Wissensverarbeitung und das Aufgreifen neuer Themen wie Situiertheit, Verteilte KI und Neuronale Netzwerke gekennzeichnet ist.

Symbolische Repräsentation – die Wissensebene

In allen Entwicklungsphasen der KI wurde mit jeweils verschiedenen Ansätzen das Ziel verfolgt, Prinzipien der Informationsverarbeitung zu erforschen und zwar dadurch, daß

1. strikte *Formalisierungen* versucht und

2. exemplarische Realisierungen durch *Implementation* vorgenommen werden.

Dabei galt und gilt auch heute noch zentrale Aufmerksamkeit der Repräsentation und Verarbeitung von Symbolen als wichtige Basis interner Prozesse, von denen man annimmt, daß sie rationales Denken konstituieren. In der Arbeit an ihrem „Logic Theorist“ hatten Simon und Newell erste Eindrücke von den Möglichkeiten des Computers zur Verarbeitung nicht-numerischer Symbole erlangt. Symbole wurden dabei als bezeichnende Objekte verstanden, die den Zugriff auf Bedeutungen – Benennungen und Beschreibungen – ermöglichen. Die symbolische Ebene, repräsentiert in den frühen Arbeiten von Newell, Shaw und Simon [Newell et al. 1958] wie auch 1956 von Bruner, Goodnow und Austin (vgl. [Bruner et al. 1956]), ermöglicht die Betrachtung von Plänen, Prozeduren und Strategien; sie stützt sich ebenfalls auf Vorstellungen regelgeleiteter generativer Systeme [Chomsky 1957].

Dabei ist der wichtigste Aspekt, daß sich geistige Fähigkeiten des Menschen auf der symbolischen Ebene unabhängig von der Betrachtung neuronaler Architekturen und Prozesse untersuchen lassen.³ Gegenstand der symbolischen KI sind folglich nicht das Gehirn und Prozesse des Abrufs von Gedächtnisbesitz, sondern vielmehr die Bedeutung, die sich einem Prozeß vermöge symbolischer Beschreibungen zuordnen läßt. Unbestreitbar hatten die Arbeiten von Newell und Simon in der Präzisierung des Informationsverarbeitungs-Paradigmas einen entscheidenden forschungsorientierenden Einfluß, der zur Ausformung der „symbolischen KI“ führte.

Ein zentrales Paradigma der symbolischen KI wurde mit der Beschreibung des „*intelligenten Agenten*“ („general intelligent agent“) [Newell, Simon 1972] formuliert. Auf einer abstrakten Ebene betrachten die Autoren den Gedächtnisbesitz des Individuums und seine Fähigkeit, beim Handeln in der Welt darauf aufzubauen, als funktionale Qualität, die sie mit *Wissen* bezeichnen. Der intelligente Agent verfügt über Sensoren, zur Wahrnehmung von Information aus seiner Umgebung, und über Aktuatoren, mit denen er die äußere Welt beeinflussen kann. Spezifisch für diese Auffassung ist, daß der Agent zu einem internen „Probearbeiten“ fähig ist: Bevor er in der Welt handelt und sie dadurch möglicherweise irreversibel verändert, manipuliert er eine interne Repräsentation der Außenwelt, um den Effekt alternativer, ihm zur Verfügung stehender Methoden abzuwägen. Diese sind ihm in einem internen Methodenspeicher verfügbar, und ihre Exploration wird durch ebenfalls intern verfügbares Weltwissen geleitet.

Die Fragen, mit denen sich vor allem Newell in den frühen achtziger Jahren befaßte, waren die folgenden [Newell 1981]:

- Wie kann Wissen charakterisiert werden?
- Wie steht eine solche Charakterisierung in Beziehung zur Repräsentation?
- Was genau zeichnet ein System aus, wenn es über „Wissen“ verfügt?

Die Hypothese einer *Wissensebene* („Knowledge Level Hypothesis“) wurde von Newell in seinem Hauptvortrag auf der ersten National Conference on Artificial Intelligence in Stanford 1980 (s. [Newell 1981]) unterbreitet. In ihr wird eine besondere Systemebene postuliert,

³Diese These ist allerdings nicht unumstritten; vor allem von Forschern auf dem Gebiet der neuronalen Netze wurde in den letzten Jahren versucht, sie zu relativieren.

über die Ebene der Programmsymbole (und die Ebenen von Registertransfer, logischem und elektronischem Schaltkreis und physikalischem Gerät) hinausgehend, die durch Wissen als das Medium charakterisiert ist. Repräsentationen existieren auf der Symbolebene als Datenstrukturen und Prozesse, die einen Wissensbestand auf der Wissensebene realisieren. Die Verbindung zwischen Wissen und intelligentem Verhalten wird durch das *Rationalitätsprinzip* beschrieben, welches besagt: Wenn ein Agent Wissen darüber hat, daß eine seiner möglichen Aktionen zu einem seiner Ziele beiträgt, dann wird der Agent diese Aktion wählen. In dieser Perspektive spielt Wissen die Rolle der Spezifikation dessen, wozu eine Symbolstruktur in der Lage sein soll. Wichtiger noch wird mit dieser Konzeption Wissen als eine *Kompetenz* betrachtet – als ein Potential, Aktionen zu generieren (zu handeln) – und mithin als eine abstrakte Qualität, die an eine symbolische Repräsentation gebunden sein muß, um einsatzfähig zu sein. Newell und Simon postulieren, daß ein dafür geeignetes physikalisches Symbolsystem zur Ausstattung eines jeden intelligenten Agenten gehört [Newell, Simon 1972; Newell 1980].

Wissensrepräsentation und -modellierung

Eine zentrale Feststellung in Newells oben genanntem Ansatz besagt, daß Logik ein fundamentales Werkzeug für Analysen auf der Wissensebene ist und daß Implementationen von Logikformalismen als Repräsentationsmittel für Wissen dienen können. Der Wissens-ebenen-Ansatz in der KI ist damit ein Versuch der Mathematisierung bestimmter Aspekte der Intelligenz – unabhängig von Betrachtungen ihrer Realisierung auf Symbolebene; dies betrifft vor allem die Aspekte des rationalen Handelns und des logischen Schlußfolgerns beim Problemlösen. Dementsprechend werden Logikformalismen vielfach in der KI benutzt, um eine explizite Menge von Überzeugungen (für wahr gehaltene Aussagen, engl. „Beliefs“) eines rationalen Agenten zu beschreiben. Eine solche Menge von Überzeugungen, ausgedrückt in einer Repräsentationssprache, wird typischerweise mit dem Terminus *Wissensbasis* bezeichnet.

Diese logikorientierte Auffassung der Wissensebene hat zur Klärung zahlreicher Debatten, die bis Ende der siebziger Jahre um den Begriff der internen Repräsentation geführt wurden, beigetragen [Brachman, 1979]. Zum Beispiel wurden unterschiedliche Ansätze der Darstellung von Wissen – wie semantische Netzwerke oder Frame-Strukturen – als notationelle Varianten herausgestellt, soweit es Ausdrucks- und Schlußfähigkeit anbelangt [Charniak, McDermott 1985]. Die Prominenz, die diese alternativen Notationen nach wie vor in vielen Anwendungsfeldern haben, leitet sich aus ihrer „Objekt-Zentriertheit“ ab, die eine Bequemlichkeit der Beschreibung von Wissensbeständen durch ausgezeichnete Konzepte bietet, und das gesamte Gebiet der objekt-orientierten Programmierung ist in Verbindung damit großgeworden. Formalismen für die Wissensrepräsentation sind mittlerweile sehr weitgehend und grundsätzlich untersucht worden, und es zeichnen sich Bestrebungen zur Standardisierung ab. Zentrale Gesichtspunkte sind hier u.a. die Ausdrucksfähigkeit und die Komplexität von Repräsentationen, aber auch ihre prädikatenlogische Rekonstruktion bzw. Spezifikation.

Ein standardisiertes Vorgehen bei der Wissens- und Domänenmodellierung hat sich als aus-

gesprochen schwierig erwiesen. Die Erfahrung hat gezeigt, daß beim Entwurf eines wissensbasierten Systems vor allem hinsichtlich der Wissensbasis eine komplexe kreative Design-Leistung gefordert ist, die mit beinahe jedem neuen System und für jeden weiteren Gegenstandsbereich neu erbracht werden muß. Deshalb stellt sich angesichts des wachsenden Umfangs projektierter wissensbasierter Systeme immer stärker die Frage nach einer Wiederverwendbarkeit schon existierender Wissensbasen bzw. nach einer Aggregation großer Wissensbasen aus bibliotheksmäßig gesammelten oder inkrementell entwickelten Teilen. Vorstöße in dieser Richtung sind verschiedene Ansätze des „Knowledge Sharing“ (vgl. [Neches et al. 1991]) oder der Modularisierung wissensbasierter Systeme (s. z.B. [Meyer-Fujara et al. 1994]). Die Entwicklung allgemeiner Vorgehensweisen zum Entwurf wissensbasierter Systeme orientiert sich dabei an der von Newell proklamierten Wissensebene mit ihrer Abgrenzung von der Ebene der symbolischen Verarbeitung (vgl. z.B. die KADS-Methode [Wielinga et al. 1992]). Hiermit verbindet sich der Anspruch, die Wissensinhalte und ihre Funktion für einen Systemzweck ins Zentrum der Modellierungstätigkeit zu stellen und zu abstrahieren von der Form der symbolischen Darstellung des Wissens und den symbolverarbeitenden Prozeduren, die die Funktionalität eines Systems hervorbringen.

Vorangetrieben wurde diese Entwicklung vornehmlich im Kontext des Entwurfs von Expertensystemen. Typischerweise ist der Gegenstandsbereich hier ein eng umrissenes Spezialgebiet, in dem hohes Potential an spezifischer Problemlösefähigkeit in einem weitgehend vorab festgelegten Verwendungsrahmen verlangt ist. Bei der Entwicklung von Systemen, die zur semantischen Verarbeitung von natürlicher Sprache fähig sind, geht es dagegen zentral um die Identifikation und Modellierung intersubjektivierbarer Bestände an Welt- oder Hintergrundwissen. Die Modellierung von Alltagswissen, d.h. von allgemeinen Kenntnissen und Fertigkeiten, erhält einen wesentlich höheren Stellenwert und muß umfangreicheren Begriffsuniversen Rechnung tragen. Bereits bei Expertensystemen war nun aber eine bittere Erfahrung, daß maschinell verfügbare Expertise genau dort ihre Grenzen hat, wo Alltagswissen und Alltagserfahrung entscheidend zum Tragen kommen. Menschliches Problemlösen zeichnet sich dadurch aus, daß das dabei verwendete Wissen zumeist vage und unvollständig ist. Die Qualität menschlicher Experten zeigt sich gerade darin, daß und wie sie unerwartete Effekte und Ausnahmesituationen aufgrund ihrer Berufserfahrung bewältigen können, daß sie aus Erfahrung *lernen*, ihr Wissen also ständig erweitern, und daß sie aus allgemeinem Wissen nicht nur nach festen Schlußregeln, sondern auch durch Analogie und mit Intuition Folgerungen gewinnen.

Richten sich die systematischen Ansätze im Bereich Expertensysteme vornehmlich auf Strukturen und Typentaxonomien von Problemlösungsaufgaben („Problemlöseontologien“), so steht die systematische Untersuchung formal repräsentierbarer kognitiver Modelle menschlicher Weltwahrnehmung, wie sie etwa in den Projekten CYC [Lenat, Guha 1990] und LILOG (s. hierzu [Klose et al. 1992]) angegangen wurde, erst am Anfang. Hier geht es nicht nur darum, generische Problemlöseaufgaben zu betrachten, sondern auch darum, das Fakten- und Relationengefüge diverser Domänen wie auch der Strukturen von Wissensmodellen des Menschen zu erschließen – z.B. durch gestaffelte generische und bereichsbezogene „Ontologien“ –, um den Entwurf wissensbasierter Systeme bei der Wissensrepräsentation zu systematisieren.

Verteilung und Situiertheit

Eine vom Modell des „general intelligent agent“ abweichende Perspektive wird in Minskys „Society of Mind“ [Minsky 1986] eingenommen. Dieses Paradigma, welches intelligentes Verhalten in der verteilten Tätigkeit vieler kleiner und noch kleinerer Systeme („Agenten“) begründet sieht, beeinflusst eine wachsende Zahl von Forschern in der KI und führt offensichtlich zu einer andersartigen Vorstellung von Intelligenz. Auf der technischen Seite haben die Versuche, immer größer und komplexer werdende wissensbasierte Systeme zu entwickeln, Nachteile zentralisierter „single-agent“-Architekturen enthüllt und die Konzipierung einer „Verteilten Künstlichen Intelligenz“ (VKI) beflügelt [Adler et al. 1992; Müller 1993]. Sog. Multi-Agenten-Systeme stellen den Aspekt der aufgabenbezogenen Kooperation unabhängiger Teilsysteme (autonomer Agenten) heraus, wobei kein Agent eine globale Sicht des gesamten Problemlösungsprozesses innehat, also keine zentrale Systemsteuerung vorliegt.

Eines der kritischsten Probleme in bisherigen Intelligenzmodellen der KI wie auch in vielen technischen Anwendungen liegt darin, daß das benötigte Weltwissen kaum jemals vollständig verfügbar bzw. modellierbar ist. Dies beruht auf der kontextuellen Variabilität und der Vielzahl von Situationen, mit denen ein intelligenter Agent konfrontiert sein wird. Die neue Forschungsrichtung der „situierten KI“ geht von der Erkenntnis aus, daß die Handlungsfähigkeit eines intelligenten Agenten entscheidend von seiner Verankerung in der aktuellen Situation abhängt [Brooks 1991]. *Situiertheit* bezieht sich auf die Fähigkeit eines intelligenten Systems, die aktuelle Situation – durch Wahrnehmung seiner Umgebung oder durch Kommunikation mit kooperierenden Partnern – in weitestgehendem Maße als Informationsquelle auszunutzen, um auch Situationen bewältigen zu können, für die kein komplettes Weltmodell vorliegt (vgl. [Lobin 1993]).

Neuronale Netzwerke

Bereits 1949, als die ersten Digitalrechner ihren Siegeszug angetreten hatten, wurde von D.O. Hebb die Grundlage für ein Verarbeitungsmodell formuliert, das eher in der Tradition des Analogrechnens steht [Hebb 1949]. Er postulierte, daß eine Menge von (formalen) Neuronen dadurch lernen könnte, daß bei gleichzeitiger Reizung zweier Neuronen die Stärke ihrer Verbindung vergrößert würde. F. Rosenblatt griff diese Idee auf und arbeitete sie zu einer Alternative zum Konzept der KI in symbolverarbeitenden Maschinen aus:

„Viele der Modelle, die diskutiert wurden, beschäftigen sich mit der Frage, welche logische Struktur ein System besitzen muß, um eine Eigenschaft X darzustellen. . . Ein alternativer Weg, auf diese Frage zu schauen, ist folgender: Was für ein System kann die Eigenschaft X (im Sinne einer Evolution) hervorbringen? Ich glaube, wir können in einer Zahl von interessanten Fällen zeigen, daß die zweite Frage gelöst werden kann, ohne die Antwort zur ersten zu kennen.“ [Rosenblatt 1962]

1956, im selben Jahr, als Newell und Simons Programm einfache Puzzles lösen und Sätze der Aussagenlogik beweisen konnte, war Rosenblatt bereits in der Lage, ein künstliches neuronales Netzwerk, das Perceptron, lernen zu lassen, gewisse Arten ähnlicher Muster zu

klassifizieren und unähnliche auszusondern. Er sah darin eine gewisse Überlegenheit seines Ansatzes und stellte fest:

„Als Konzept, so scheint es, hat das Perceptron ohne Zweifel Durchführbarkeit und Prinzip nichtmenschlicher Systeme begründet, die menschliche kognitive Funktionen darstellen können... Die Zukunft der Informationsverarbeitungssysteme, die mit statistischen eher als logischen Prinzipien arbeiten, scheint deutlich erkennbar.“ [Rosenblatt 1962]

Zunächst jedoch gewann der symbolische Ansatz in der KI die Oberhand, was nicht zuletzt darin begründet war, daß Rosenblatts Perceptron gewisse einfache logische Aufgaben nicht lösen konnte – eine Beschränkung, die aber ohne weiteres überwunden werden kann. So erfuhr Rosenblatt in den letzten zehn Jahren eine Rehabilitation, und das Arbeitsgebiet der Neuronalen Netze bzw. des „Konnektionismus“ hat sich rapide zu einem umfangreichen Teilgebiet der KI entwickelt. Der theoretische Informatiker B. Mahr hat diese Konzeption treffend charakterisiert, so daß wir hier auf seine Darstellung zurückgreifen [Mahr 1989]:

„Für die Erzeugung künstlicher Intelligenz wird ein Maschinenmodell zugrundegelegt, das Arbeitsweise und Struktur des Neuronengeflechts im Gehirn imitiert. Den Neuronenkernen mit ihren Dendriten und deren Verknüpfung über Synapsen entsprechen 'processor'-Knoten, die über Verbindungen miteinander gekoppelt sind. ... Die Idee des Lernens durch die Stärkung der Verbindung, die auch schon Rosenblatts Perceptron zugrundelag, findet sich hier in der Fähigkeit wieder, daß die Gewichte der Verbindungen sich ändern können und daß so nicht nur das Pattern der Verbindungen wechselt, sondern auch das Verhalten des gesamten Systems. ... Das 'Wissen', das in einem System steckt, erscheint dann als Pattern der Verbindungsgewichte. ... Künstliche neuronale Netze geben als Computerarchitektur die Manipulation bedeutungstragender Symbole auf ... Sie stellen 'Wissen' ... nicht als aus einzelnen Wissensbestandteilen zusammengesetztes Ganzes dar.“

Als Vorteile künstlicher neuronaler Netze gelten ihre Eigenschaften der verteilten Repräsentation, der Darstellung und Verarbeitung von Unschärfe, der hochgradig parallelen und verteilten Aktion und die daraus resultierende Geschwindigkeit und hohe Fehlertoleranz. Dennoch gilt für beide Ansätze, den subsymbolischen und den symbolischen, daß wohl keiner von ihnen alleine die Methodik der KI ausmachen kann. Vielmehr wird die Zukunft in einer Synthese beider liegen, in hybriden Systemen, in die jeder Ansatz seine besonderen Stärken einbringen kann.

Teilbereiche der KI

Mit den folgenden Kapiteln wird versucht, die etablierten Grundlagen- und Anwendungsbereiche der KI so weit wie möglich abzudecken:

- Die *Wissensrepräsentation* befaßt sich mit der Darstellung von Objekten, Ereignissen und Verläufen und von Performanz- und Meta-Wissen durch formale, i.a. logikbasierte Systeme.
- Kalküle für *automatisches Beweisen* werden u.a. auf die Herstellung und Überprüfung mathematischer Beweise sowie die Analyse (Verifikation) und Synthese von Program-

men mit deduktiven Methoden angewandt.

- Verfahren des *maschinellen Lernens* betreffen Programmsysteme, die aus „Erfahrung“ lernen, also neues Tatsachen- und Regelwissen gewinnen oder Priorisierungen adaptieren können.
- Die Betrachtung von *Kognition* als Informationsverarbeitung liefert Grundlagen für eine Fülle von Methoden, u.a. für *heuristische Suchverfahren*, die dem Zweck dienen, in hochkomplexen Suchräumen schnellere Lösungswege zu finden.
- Verfahren zur *Verarbeitung der natürlichen Sprache* richten sich darauf, Einsicht in den „Mechanismus“ der Sprache – ihren Aufbau, ihre Verarbeitung und ihre Verwendung – zu gewinnen und diese für die Mensch-Maschine-Interaktion nutzbar zu machen.
- Beim *Bildverstehen* geht es um Aufgaben der Wahrnehmung, um Merkmale aus optischen Daten zu gewinnen und daraus Interpretationen von stehenden und bewegten *Bildern* zu erzeugen.
- Zu den zentralen Aufgaben vieler KI-Systeme gehört die Lösung komplexer *Planungs-, Konfigurations-, und Diagnoseprobleme*, die an Vorgehensweisen menschlicher Experten orientiert werden.
- *Künstliche Neuronale Netze* betrachten Verarbeitungsmodelle, die sich durch Lernfähigkeit, Darstellung und Verarbeitung von Unschärfe, hochgradig parallele Aktion und Fehlertoleranz auszeichnen.
- In der *KI-Programmierung* realisiert man oft neuartige Verarbeitungsmodelle und Kontrollstrukturen, die u.a. durch abstrakte Systemarchitekturen mit verteilter Kontrolle erforderlich werden. Hieraus resultiert eine reichhaltige Palette von speziellen Programmiersprachen und -systemen.

Aus der Grundlagenforschung ging eine Reihe zunächst eher prototypischer Anwendungssysteme hervor, viele ihrer Ergebnisse sind aber heute bereits Teil in der Praxis genutzter Anwendungen geworden. Die Vorteile der *KI-Technologie* sind im wesentlichen von zweierlei Art: Zum einen eröffnet sie neue Anwendungen wie z.B. im maschinellen Sprach- oder Bildverstehen, in der Robotik und mit Expertensystemen. Zum anderen aber ermöglicht sie auch bessere Lösungen für alte Anwendungen; hierzu gehören vor allem die maschinelle Unterstützung von Planen, Entscheiden und Klassifizieren sowie die Verwaltung, Erschließung und Auswertung großer Wissensbestände und schließlich die Simulation und die Steuerung technischer Anlagen.

Gegenwärtig zeichnet sich im Gebiet „Künstliche Intelligenz“ ein Paradigmenwechsel ab; es bewegt sich von einer globalen Betrachtung intelligenten Verhaltens hin zu einer Sicht von einfacheren interagierenden Systemen mit unterschiedlichen Repräsentationen – oder auch gar keiner Repräsentation –, wie die Arbeiten zu Multi-Agenten-Systemen, Verteilter KI und Neuronalen Netzwerken zeigen. Diese Ansätze bringen eine Erweiterung auf die Untersuchung „situierter“ Systeme ein, welche durch Sensoren und Aktuatoren in ständigem Austausch mit ihrer Umgebung stehen, um etwa auch während einer Problemlösung Situationsdaten aufzunehmen und auszuwerten. Bisher hat kein einzelner Ansatz eine Perspektive geboten, mit der sich alle Aspekte intelligenten Verhaltens reproduzieren oder erklären ließen, wie es auf der Dartmouth-Konferenz als Programm formuliert wurde. Ei-

ne kürzliche Ausgabe des „Scientific American“⁴ zitiert Minsky mit dem treffenden Satz: „The mind is a tractor-trailor, rolling on many wheels, but AI workers keep designing uni-cycles.“ Erscheinen nach wie vor noch viele Fragen als grundsätzlich ungelöst, so gibt es doch Evidenz dafür, daß mittlerweile mehr als ein „Rad“ untersucht wird und daß gerade die Integration verschiedener Ansätze weitergehende Perspektiven für die Grundlagenforschung und Anwendungsentwicklung eröffnet.

Danksagung. Die Autoren danken Clemens Beckstein für eine Reihe hilfreicher Hinweise.

Literaturverzeichnis

- [Adler et al. 1992] Adler, M., Durfee, E., Huhns, M., Punch, W., Simoudis, E.: *AAAI Workshop on Cooperation Among Heterogeneous Intelligent Agents*. AI Magazine 13 (2), 1992, 39–42.
- [Brachman 1979] Brachman, R.J.: *On the Epistemological Status of Semantic Networks*. In: Findler, N.V. (Ed.): *Associative Networks: Representation and Use of Knowledge by Computers*. New York: Academic Press, 1979, 3–50.
- [Brooks 1991] Brooks, R.A.: *Intelligence without reason*. Proceedings IJCAI-91, Sydney, 1991, 569–595.
- [Bruner et al. 1956] Bruner, J.S., Goodnow, J.J., Austin, G.A.: *A Study of Thinking*. New York: Wiley, 1979.
- [Charniak, McDermott 1985] Charniak, E., McDermott, D.: *Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley, 1985.
- [Chomsky 1957] Chomsky, N.: *Syntactic Structures*. The Hague: Mouton, 1957.
- [Feigenbaum, Feldman 1963] Feigenbaum, E.A., Feldman, J.: *Computers and Thought*. New York: McGraw-Hill, 1963.
- [Gardner 1985] Gardner, H.: *The Mind's New Science — A History of the Cognitive Revolution*. New York: Basic Books, 1985.
- [Hebb 1949] Hebb, D.O.: *The Organization of Behavior*. New York: Wiley, 1949.
- [Irrgang, Klawitter 1990] Irrgang, B., Klawitter, J.: *Künstliche Intelligenz – Technologischer Traum oder gesellschaftliches Trauma?* In: Irrgang, B., Klawitter, J. (Hg.): *Künstliche Intelligenz*. Edition Universitas, Stuttgart: Hirzel, 1990, 7–54.
- [Klose et al. 1992] Klose, G., Lange, E., Pirlein, Th. (Hg.): *Ontologie und Axiomatik von LILOG*. Berlin: Springer (IFB 307), 1992.
- [Lenat, Guha 1990] Lenat, D.B., Guha, R.V.: *Building Large Knowledge-Based Systems — Representation and Inference in the Cyc Project*. Reading, Ma: Addison-Wesley, 1990.
- [Lobin 1993] Lobin, H.: *Situiertheit*. KI, 1993, Nr. 1, (Rubrik KI-Lexikon), 63
- [McCorduck 1979] McCorduck, P.: *Machines Who Think*. San Francisco: Freeman, 1979.

⁴Scientific American, Nov. 1993, Profiles: „Marvin L. Minsky – The Mastermind of Artificial Intelligence“, S. 14–15

- [Mahr 1989] Mahr, B.: *Chaos-Connection. Einwände eines Informatikers*. Kursbuch 98, 1989, 83-99.
- [Meyer-Fujara et al. 1994] Meyer-Fujara, J., Heller, B., Schlegelmilch, S., Wachsmuth, I.: *Knowledge-level modularization of a complex knowledge base*. In: Nebel, B., Dreschler-Fischer, L. (Eds.): *KI-94: Advances in Artificial Intelligence*. Berlin: Springer (LNAI), 1994, 214-225.
- [Minsky 1986] Minsky, M.L.: *The Society of Mind*. New York: Simon & Schuster, 1986.
- [Müller 1993] Müller, J. (Hg.): *Verteilte Künstliche Intelligenz – Methoden und Anwendungen*. Mannheim: BI Wissenschaftsverlag, 1993.
- [Neches et al. 1991] Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, W.: *Enabling Technology for Knowledge Sharing*. AI Magazine 12(3), 1991, 37-56.
- [Newell, Simon 1956] Newell, A., Simon, H.A.: *The Logic Theory Machine*. IRE Transactions on Information Theory, September 1956. Abgedruckt in: [Feigenbaum, Feldman 1963].
- [Newell et al. 1958] Newell, A., Shaw, J.C., Simon, H.A.: *Chess playing programs and the problem of complexity*. IBM Journal of Research and Development 2(4), 1958. Abgedruckt in: [Feigenbaum, Feldman 1963].
- [Newell, Simon 1972] Newell, A., Simon, H.A.: *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.
- [Newell 1980] Newell, A.: *Physical Symbol Systems*. Cognitive Science 4, 1980, 135-183
- [Newell 1981] Newell, A.: *The Knowledge Level*. AI Magazine 2(2), 1981, 1-20. Wieder veröffentlicht in Artificial Intelligence 18(1), 1-20.
- [Rosenblatt 1962] Rosenblatt, F.: *Strategic Approaches to the Study of Brain Models*. In: v.Foerster, H. (Ed.): *Principles of Self-Organization*. Elmsford, N.Y.: Pergamon Press, 1962, 387.
- [Shannon, McCarthy 1956] Shannon, C.E., McCarthy, J.: *Automata Studies*. Annals of Mathematics Studies No. 34. Princeton, NJ: Princeton University Press, 1956.
- [Wielinga et al. 1992] Wielinga, B.J., Schreiber, A.Th., Breuker, J.A.: *KADS: A Modelling Approach to Knowledge Engineering*. Knowledge Akquisition 4, 1992, 5-53.
- [Winston 1992] Winston, P.H.: *Artificial Intelligence*. (3rd edition) Reading, MA: Addison-Wesley, 1992.