

# 1. Künstliche Intelligenz – Grenzen und Möglichkeiten

## 1.1. Begriffsbestimmung Intelligenz – künstliche Intelligenz auf der Basis des symbolverarbeitenden Ansatzes

AHRWEILER unterscheidet zwei Bearbeitungsstrategien in der KI-Forschung. Den symbolverarbeitenden Ansatz und den konnektionistischen Ansatz. Diese unterschiedlichen Ansätze haben „für kognitionswissenschaftliche Fragestellungen einerseits und ... verhandelte Stellungnahmen zu erkenntnistheoretischen Positionen andererseits...“ (Ahrweiler 1995: S.22) verschiedenartige Auswirkungen. Deshalb wurden die Ausführungen ausdrücklich auf den symbolverarbeitenden Ansatz reduziert, um einer Verflechtung zweier verschiedener Theorien aus dem Weg zu gehen.

### 1.1.1. Was ist Intelligenz – Eine Dimensionierung

Intelligenz ist „die Fähigkeit zu problemlösendem, einsichtigem Verhalten.“ (Cruse/Dean/Ritter, 1998: S. 9) Der Brockhaus definiert Intelligenz als Klugheit, Fähigkeit der Auffassungsgabe, des Begreifens, Urteilens; geistige Anpassungsfähigkeit an neue Aufgaben. Während Meyers enzyklopädisches Lexikon Intelligenz als die Bewältigung neuartiger Situationen durch problemlösendes Verhalten versteht. SPÖRL umschreibt den Begriff der Intelligenz wie folgt: „Mit Intelligenz meine ich das Vermögen eines Lebewesens oder eines Apparates, Informationen von außen, Beobachtungen, Erfahrungen zu ordnen, Zusammengehörigkeiten zu entdecken, die Informationen damit auszuwerten, das alles, um auf diese Weise zu abstrahieren, um sie miteinander verknüpfen zu können.“ Bei diesen vielen unterschiedlichen Definitionsansätzen lässt sich jedoch eine Gemeinsamkeit feststellen: Die meisten Definitionen setzten das Vorhandensein einer Ratio voraus, in dieser erschöpft sich jedoch bei weitem nicht der Begriff der Intelligenz. DREYFUS und DREYFUS bemerken hierzu: „Zur Intelligenz gehört mehr als nur kalkulierbarer Verstand.“ (Dreyfus/Dreyfus, 1988: S.61) Sie unterscheiden irrationales, rationales und arationales Handeln. Wobei ersteres im Widerspruch zur Logik und zum Verstand steht, die Rationalität der Vernunft folgt und Arationalität „nennen wir ... Handlungen, die ohne bewusstes, analytisches Zerlegen und Rekombinieren auskommen.“ (Dreyfus/Dreyfus, 1988: S.62)

Da der Begriff der Intelligenz so viele unterschiedliche Facetten menschlicher Eigenschaften anspricht, haben sich in den vergangenen Jahrzehnten mehrere Forscher auf eine Dimensionierung von Intelligenz eingelassen. Einige dieser Versuche sollen an dieser Stelle wiedergegeben werden, um den Begriff in möglichst breiter Form zu charakterisieren.

Bei CRUSE, DEAN und RITTER wird die Frage gestellt, ob „Intelligenz eine einheitliche, unteilbare Fähigkeit“ ist, oder ob „sie die Summe vieler Einzelfähigkeiten“ (Cruse/Dean/Ritter, 1998: S.16) darstellt. Die Erkenntnisse der Neurophysiologie hat den Forschern, die sich mit Intelligenz beschäftigen, diese Trennung schon fast in den Mund gelegt, sie gehen nämlich davon aus, dass unterschiedlichen Bereichen unseres Gehirns unterschiedliche Aufgaben zukommen. Auch die drei oben genannten Autoren kommen zu dem Schluss, den Begriff Intelligenz zu dimensionieren:

1. „Ein intelligentes System<sup>1</sup> sollte autonom sein (wörtlich: sich selbst das Gesetz, die Regel gebend)...“. Das System sollte hierbei sein „Verhalten weitestgehend selbst bestimmen.“ (Cruse/Dean/Ritter, 1998: S.23) Dies erweist sich jedoch in vielen Fällen als schwer trennbar, ob das System sich selbst bestimmt, oder ob es von außen bestimmt wird. Dennoch sollte es prinzipiell eine Unterscheidung zwischen eigen- und fremdbestimmtem Verhalten treffen können.
2. „Ein intelligentes System sollte Intentionen besitzen.“ (Cruse/Dean/Ritter, 1998: S.23) Es sollte also in der Lage sein, „sich selbst die Ziele seines Verhaltens auszuwählen.“ (Cruse/Dean/Ritter, 1998: S.23) Auch wenn, wie schon oben beschrieben, das Wort „selbst“ einige Probleme aufwirft.
3. „Ein intelligentes System sollte sich anpassen und aus Erfahrungen lernen können.“ (Cruse/Dean/Ritter, 1998: S.24) Es sollte auf die Veränderungen seiner Umwelt reagieren können, wobei vorausgesetzt wird, dass das System selbst in der Lage ist, sich zu verändern. Es dürfen also keine starren

---

<sup>1</sup> Des öfteren wird der Begriff des Systems im Laufe dieser Arbeit benutzt. Er wird auf die Definition von Cruse, Dean und Ritter zurückgeführt. „Es muss zum einen die Möglichkeit besitzen, Signale von außen aufzunehmen, z.B. durch Sensoren, und es muß zum zweiten Signale nach außen abgeben können.“ (Cruse/Dean/Ritter, 1998: S.19)

Strukturen vorliegen, sondern es muss in einem dialogischen Verhältnis zu seiner Umwelt stehen. In einem ähnlichen Zusammenhang schreiben DEAN, HOLK und RITTER, dass ein intelligentes „System ... offen sein“ (Cruse/Dean/Ritter, 1998: S.27) sollte. Es darf also nicht nur in einem begrenzten Raum, z.B. einem Labor überlebensfähig sein, sondern muss auch mit unbekanntem Situationen zurecht kommen.

4. „Eine wichtige Eigenschaft von Intelligenz besteht auch darin, den Erfolg eines Verhaltens beurteilen zu können.“ (Cruse/Dean/Ritter, 1998: S.24) Die drei Autoren machen dies am Beispiel der Fliege deutlich, die selbst nach mehreren Versuchen immer wieder gegen die Fensterscheibe fliegt und damit den Misserfolg ihres Verhaltens nicht reflektieren kann. Ein intelligentes System sollte zumindest nach dem Prinzip von „Trial and error“ zu einer alternativen Problemlösung gelangen, um „Sackgassen“ zu vermeiden, die wie im Beispiel der Fliege lebensbedrohlich enden können.
5. „Eine weitere wichtige Eigenschaft ist die Fähigkeit zur Generalisierung.“ (Cruse/Dean/Ritter, 1998: S.24) Der Begriff der Generalisierung wird hier in zweifacher Hinsicht gebraucht. Zum einen sollte das System über eine gewisse „Fehlertoleranz“ verfügen, die es ermöglicht, so noch nicht wahrgenommene Dinge einzugliedern. Generalisierung bedeutet also hier Ähnlichkeiten zu erkennen, um sie in schon Bekanntes einzufügen. Auf der anderen Seite sollte ein intelligentes System die Fähigkeit der Generalisierung dazu benutzen können, Kategorien zu bilden, um Abstraktion zu ermöglichen. Ein Beispiel soll den Unterschied der verschiedenartigen Anforderungen an Generalisierung näher erläutern:  
Durch das Erkennen unterschiedlicher Farben ist der Mensch in der Lage Rottöne und Blautöne zu kategorisieren. Er muss aber auf der anderen Seite auch in der Lage sein, ein gesehenes Orange als einen modifizierten Rotton zu erkennen.
6. Eine andere Eigenschaft von Intelligenz „ist die Fähigkeit, zwischen Alternativen entscheiden zu können.“ (Cruse/Dean/Ritter, 1998: S.25) Meistens fällt die Wahl zwischen zwei gleichwertigen Alternativen sehr schwer. Aus diesem Grund setzen intelligente Systeme sogenannte Kostenfunktionen oder allgemeine Regeln ein, um eine Entscheidung treffen zu können, z.B.

immer der kürzeste Weg oder die Lösung, „die die wenigste Energie oder die geringste Zeit verbraucht.“ (Cruse/Dean/Ritter, 1998: S.26)

7. „Die vielleicht wichtigste Bedingung für das Auftreten von Intelligenz besteht in der Fähigkeit, Änderungen der Umwelt, z.B. als Folge eigener Aktivität vorhersagen zu können.“ (Cruse/Dean/Ritter, 1998: S.27) Hierbei unterstellen die drei Autoren dem System schon so etwas wie Bewusstsein, weil es sich um ein selbstreferentielles System handelt, das was MATURANA und VARELA als Autopoieses beschreiben.

### **1.1.2. Auswertung und Definition**

„Ein System ist also dann intelligent, wenn es in einer gegebenen und sich verändernden Umwelt die Chancen seiner Selbsterhaltung im Vergleich zu seinem aktuellen Zustand verbessern kann.“ (Cruse/Dean/Ritter, 1998: S.27)

CRUSE, DEAN und RITTER haben sich in ihrer abschließenden Definition auf eine eher biologische Sichtweise gestützt. So wird das Ziel der Selbsterhaltung zum Kriterium, das den Dimensionen von Intelligenz ihre Position zuweist und sie untereinander verbindet. Diese obengenannte Definition sollte also in einem Zusammenhang zu der Dimensionierung gesehen werden, um eine für diese Arbeit zufriedenstellende Begriffsbestimmung für den Terminus „Intelligenz“ festzulegen.

CRUSE, DEAN und RITTER haben sich bemüht, den Begriff Intelligenz immer auf Systeme zu beziehen, um zu verdeutlichen, dass diese Eigenschaft nicht nur dem Menschen zukommt, sondern durchaus auch Tieren, wenn nicht sogar Maschinen. Für eine Arbeit über künstliche Intelligenz ist solch ein Definitionsansatz förderlich, da dadurch nicht im Vorhinein ausgeschlossen wird, dass auch Digitalcomputer intelligent sein können.

### **1.1.3. Was ist künstliche Intelligenz? – Der symbolverarbeitende Ansatz**

Der symbolverarbeitende Ansatz ist der ältere von beiden oben genannten Ansätzen (vgl. Kap. 1.1.) in der KI-Forschung. Nach AHRWEILER (Ahrweiler 1995: S.22) sollen hierbei die einzelnen Denkschritte des Menschen direkt „in die Sprache eines Computerprogramms übersetzt“ werden. Nach NEWELL und

SIMON sind „kognitive Prozesse Transformation von Symbolstrukturen. Symbolstrukturen wiederum sind aus elementaren Symbolen als bedeutungstragende Einheiten gemäß syntaktischer Regeln zusammengesetzt.“ In diesem Zusammenhang sei die These von NEWELL und SIMON eine hinreichende Voraussetzung von intelligentem Verhalten (Strube, 1993: S.304). „Die Fähigkeiten, die diese Leistungen [z.B. Verknüpfung von Symbolketten] ermöglichen“, wollen wir „auf der funktionellen Ebene mit rationaler Intelligenz bezeichnen.“ (Cruse/Dean/ Ritter, 1998: S.10) Der Begriff der rationalen Intelligenz übernimmt auch für die Definition der künstlichen Intelligenz eine wichtige Funktion. Digitalcomputer sind symbolverarbeitende Systeme, die den Gesetzen der Logik folgen, die wiederum auf der Ratio fußen. An dieser sehr eingeschränkten Definition von künstlicher Intelligenz liegt das Problem. Ist ein System wirklich als intelligent zu bezeichnen, wenn es sich nur auf künstliche Problemstellungen beschränkt? Diese Frage jedoch soll Gegenstand der Diskussion über Grenzen und Möglichkeiten künstlicher Intelligenz sein. (vgl. Kap. 1.2./1.3.)

Ein großer Bereich der KI-Forschung ist die Cognitive Science, die den menschlichen Geist und die Materie miteinander verbindet. Sie beschäftigt sich mit Fragen wie „Gibt es Kategorien, die menschliches und maschinelles Denken a priori beschränken?“ oder „Wie funktioniert die biologische Informationsverarbeitung?“ (Bibel/u.a., 1987: S. 5) Es sei hier nur beispielhaft auf die Ausführungen zu den Grenzen der KI hingewiesen, denn auch damit beschäftigen sich die Cognitive Science Forscher.

Die Ziele der KI-Forschung sind einerseits „to construct computer programs which exhibit behavior that we call “intelligent behavior” when we observe it in human beings“, andererseits will man durch die Entwicklung von Simulationsmodellen natürliche Kognitionsprozesse besser verstehen lernen. In diesen Definitionsansatz, künstliche Intelligenz möge menschliche Intelligenz simulieren, reiht sich auch der nun folgende Turing-Test ein, der dann als erfolgreich bezeichnet werden darf, wenn ein Mensch während einer Konversation den Unterschied zwischen einer Maschine und einem Menschen nicht mehr wahrnehmen kann.

#### 1.1.4. Der Turing – Test

Die Ausgangsfrage, die Turing gestellt hatte war, ob Maschinen denken können, bzw. in der Lage sind einen Menschen erfolgreich zu imitieren. Im Folgenden soll sein Experiment beschrieben werden, um daran eine kritische Diskussion anzuschließen.

„Der Gedanke besteht darin, einen Menschen vor ein einfaches Computerterminal zu setzen, so dass er mit Hilfe von Tastatur und Bildschirm Fragen an die Maschine stellen und Antworten von ihr erhalten kann. Der Versuchsperson wird dabei mitgeteilt, dass ihr Terminal in verschiedenen Sitzungen entweder mit einem Computer oder mit einem Terminal verbunden ist, das gleichfalls von einem Menschen bedient wird. Die Aufgabe des Beobachters besteht darin, im Laufe dieser Unterhaltung herauszufinden, ob sein jeweiliges Gegenüber nun ein Mensch oder eine Maschine ist. Turing fordert, dass ein künstliches System dann als wirklich intelligent bezeichnet werden muss, wenn der menschliche Beobachter diesen Unterschied nicht mehr eindeutig zu treffen vermag.“ (Cruse/Dean/Ritter, 1998: S.13)

Die Beantwortung der Frage, ob durch dieses Experiment erfolgreich der Nachweis geführt wurde, dass Maschinen denken können, verbleibt der sich nun anschließenden Diskussion.

Turing setzt durch die Formulierung seines Experimentes voraus, dass Intelligenz nur über das Verhalten messbar ist. Wobei der Begriff messbar hier eher die Bedeutung von empfindbar hat. So wird das Experiment nicht eindeutiger, sondern eher beliebiger. Der Anspruch, der diesem Test zugrunde liegt, ist jedoch nicht zu vernachlässigen. „Nicht das Verhalten an sich, sondern Verhalten im Kontext einer beliebigen Umwelt bestimmen die Bewertung der Intelligenz.“ (Cruse/Dean/Ritter, 1998: S.21) Die „beliebigen Umwelten“ sind in diesem Versuch die verschiedenen Gesprächsthemen, zu denen sich der Computer äußern kann und auf der anderen Seite die unterschiedlichen Versuchspersonen. Die von CRUSE, DEAN und RITTER formulierten Anforderungen an ein intel-

ligentes System sind also erfüllt. Wie aber die folgenden Ausführungen zeigen werden, könnte der Turing-Test mit sehr zweifelhaften Tricks bestückt werden, um die Maschine intelligent erscheinen zu lassen.

Das erste sich stellende Problem ist die Voreingenommenheit des Menschen gegenüber einem Computer. Während die Versuchsperson einem Computer vorzugsweise unterstellt, rational zu handeln, unterstellt sie einer menschlichen Interaktion vielmehr Emotionalität. Eine falsche Antwort wird also eher mit den Verhaltensweisen eines Menschen verknüpft, als mit denen eines Computers. So ist man mit der einfachen Einführung eines Zufallgenerators in der Lage, auch dem Computer eine falsche Antwort zu entlocken, die man eigentlich nur von einem Menschen erwarten würde. Der Einsatz eines solchen Zufallgenerators jedoch birgt in keiner Weise intelligentes Verhalten. (vgl. Cruse/Dean/Ritter, 1998: S.13ff.)

TURING war eine klare Trennung zwischen physischen und geistigen Tätigkeiten sehr wichtig. Deshalb bevorzugte er in seinem Experiment den schriftlichen Dialog. Das Ziel, menschliche Verhaltensweisen nachzuahmen, wäre in diesem Gedankenexperiment gelungen, dennoch ist dadurch nicht bewiesen, dass Maschinen denken können. Unterstützen möchte ich diese Behauptung mit den Ausführungen von DREYFUS und DREYFUS über das von WEIZENBAUM entwickelte Interaktionsprogramm ELIZA. Das Programm war in der Lage ähnlich einem Non-direktiven Therapeuten, die Aussagen seines Gegenübers zu spiegeln bzw. in eine Frage umzuwandeln. Die Menschen, die versuchsweise mit diesem Programm interagierten, hatten wie in TURINGS Gedankenexperiment das Gefühl, mit einem realen Therapeuten zu „sprechen“. DREYFUS und DREYFUS machten jedoch mit diesem Programm folgende Erfahrung: „...als wir eingaben „ich bin glücklich“ und uns anschließend korrigierten: „Nein, in gehobener Stimmung.“ lieferte ELIZA die Bemerkung: „Seien sie doch nicht so negativ.“ Warum das? Weil es darauf programmiert war, immer, wenn ein „Nein“ eingegeben wurde, mit diesem Tadel zu reagieren.“ (Dreyfus/Dreyfus, 1988: S.104) Das heißt, das Ziel TURINGS, einen Menschen nachzuahmen, hat ELIZA erfolgreich gemeistert, dennoch hat WEIZENBAUM kein intelligentes Programm konstruiert, denn dann hätte das Programm den Inhalt der Frage ver-

standen und hätte nicht nur auf einzelne Reizwörter mit Hilfe von Regeln reagiert. Nach KLIMSA und ISSING ist ein Computer nicht in der Lage eine Kommunikation zu führen. (vgl. Klimsa, Issing, 1995: S.347)

Insofern wird auch die Turingmaschine nicht verstehen können. Die Imitation eines Menschen ist in beiden Fällen (ELIZA/ Turingmaschine) von Außen betrachtet gelungen, die Abläufe im Inneren unterscheiden jedoch den Computer vom Menschen außerordentlich.

### **1.1.5. Kritische Betrachtung hinsichtlich einer Definition**

Was verbleibt also vom Turing-Test für eine Definition? Der Turing-Test erklärt nicht, was künstliche Intelligenz ist, sondern womit künstliche Intelligenz sich beschäftigt. Auch NEBENDAHL beschäftigt sich nicht mit der Erklärung von KI, sondern mit dessen Forschungsinhalt: „Artificial Intelligence (A.I.) is the study of how to make computers do things at which, at the moment, people are better.“ (Nebendahl, 1987: S.16) Schon an dieser Definition wird deutlich, dass künstliche Intelligenz als sehr idealistisch beschrieben wird. In einer anderen Definition gehen BIBEL u.a. sowohl auf die Vorgehensweise der KI-Forscher als auch auf die Verwendung von KI ein: „Die ‚künstliche Intelligenz‘ untersucht solche bisher dem Menschen vorbehaltenen Verhaltensweisen, indem sie sie auf dem Rechner simuliert und naturwissenschaftlicher Betrachtungsweise und damit ingenieurmäßiger Verwendung zugänglich macht.“ (Bibel/u.a., 1987: S.1)

JOHN SEARLE macht einen Unterschied zwischen der starken und der schwachen KI. So besteht bei der schwachen KI der Wert des Computers im wesentlichen in seiner Funktion als Werkzeug beim Studium des Geistes (mind). Diese These wird unterstützt von KAIL und PELLEGRINO, die die menschliche Kognition mit Hilfe einiger Begriffe der Computerwissenschaft nachzeichnen. „Dabei behandeln die meisten Kognitionspsychologen unserer Zeit die menschliche Kognition als einen Prozess der Informationsverarbeitung. Der Terminus der Informationsverarbeitung ist der Computerwissenschaft entlehnt und

gibt bereits einen deutlichen Hinweis auf die Natur des Ansatzes.“ (Kail/Pellegrino, 1988: S.54)

Bei der starken KI ist ein entsprechend programmierter Computer nicht nur ein Werkzeug zum Studium des Geistes, sondern realisiert den Geist in dem Sinne, dass von ihm behauptet werden kann, er verstehe tatsächlich und könne kognitive Zustände, wie sie bei Menschen vorzufinden sind, imitieren. (Schäfer, 1994: S.104 ff.) Während also die schwache KI nach SEARLE eher den Status einer Hilfswissenschaft für die cognitive psychology und die Philosophie hat, ist die starke KI ein eigener Forschungszweig. Das würde jedoch bedeuten, dass sich die schwache KI im Erfolgsfalle selbst auflöst bzw. deren Wissen und Methoden in andere Disziplinen eingebracht würde.

Ob starke oder schwache KI, sie ist ein übergreifendes, interdisziplinäres Forschungsgebiet unter Beteiligung der (kognitiven) Psychologie, (Computer-) Linguistik, Philosophie, Neurowissenschaften. Desweiteren lässt sich nach dieser Sammlung von Definitionen zur künstlichen Intelligenz sagen, dass die KI der Frage nachgeht, in wiefern eine Maschine in der Lage ist, menschliche Intelligenz nachzuahmen. Es fällt bei dieser Ansammlung von Aussagen über die KI schwer eine Definition zu formulieren, da sich alle Ausführungen auf die Betätigungsfelder der KI beziehen, jedoch nicht auf das Wesen der KI. Dennoch lässt sich abschließend folgende Definition über die KI für diese Arbeit zugrunde legen:

Es ist die Aufgabe der KI unter Zuhilfenahme der Kognitionspsychologie, Philosophie und der Neurowissenschaften eine Maschine so zu programmieren, dass sie intelligentes menschliches Verhalten imitieren kann.

## 1.2. Grenzen der Künstlichen Intelligenz

Wenn über die Grenzen der KI gesprochen wird, wird meist versucht, das eigene Paradigma mit allen Mitteln zu verteidigen und neue Betrachtungsweisen in das eigene Denksystem zu integrieren. AHRWEILER bezeichnet solch eine Abwehrstrategie als Synkretismus (Integration in das eigene Paradigma) hierdurch ist es möglich jede Kritik zu neutralisieren. (vgl. Ahrweiler, 1995: S.18)

Mit dem Herausstellen des Synkretismus soll analysiert werden, welche Grenzen sich die KI-Forschung wirklich steckt oder ob es sogar zur Eigenschaft der KI-Forschung gehört, ohne Grenzen auszukommen. Hierbei sollen Grenzen immer verstanden werden als unabhängig vom Entwicklungsstand nie zu erreichende Ziele.

Die Annahmen, die benannt werden, gehen auf DREYFUS und DREYFUS zurück, die mit Hilfe der direkten Anlehnung an menschliche Eigenschaften des Denkens und Handelns, die Grenzenlosigkeit der KI kritisch beleuchten wollen. In diesem Zusammenhang kommt vor allem dem oben benannten Synkretismus eine besondere Bedeutung zu, da es in der Natur dieser Annahmen liegt, menschliches Verhalten auf maschinelle Prozesse zu übertragen und integrieren zu wollen.

WAHLSTER gesteht ein, dass es „Aspekte menschlichen Verhaltens gibt, die man gar nicht mit Informationsverarbeitung erklären kann.“ (Ahrweiler 1995: S.16) Er sagt weiter, „wenn es um Anwendungen geht, bin ich immer dafür, dass man soweit möglich ganzheitlich evaluiert und biologische Aspekte, die mit der Informationsverarbeitung gar nichts zu tun haben, mitberücksichtigt. Aber unsere These, unser Forschungsparadigma, ist nun einmal, dass wir alles auf Informationsverarbeitung beziehen.“ (Ahrweiler 1995: S.16) Auch an dieser Stelle wird das Wesen des Synkretismus deutlich. Es ist unerheblich aus welchem Gebiet die erforschten Annahmen stammten, sie werden immer in das eigene Paradigma eingepasst.

### 1.2.1. Die biologische Annahme

„Das menschliche Gehirn funktioniert physiologisch wie Computerhardware.“  
(Daniel/Striebel 1993: S.23) DREYFUS ergänzt, dass „das Gehirn Informationen in einzelnen Abschnitten verarbeitet, und zwar mit Hilfe eines biologischen Äquivalents von Ein-Aus-Schaltern.“(Dreyfus 1985: S. 105/106)

DREYFUS beschreibt zwei Erkenntnisse, die diese Annahme verneinen:

1. Nach vielen neurophysiologischen Forschungsergebnissen stellt man sich das Gehirn als einen elektronischen Rechner vor. So stoßen die Neuronen im Gehirn einen elektrischen Stromstoß an benachbarte Neuronen aus, ähnlich einem Computer, der über einen elektrischen Impuls Informationen zusammenfügt. Bei einer genaueren Untersuchung der Tätigkeit eines digitalen Computers stellt man jedoch fest, dass es das Merkmal der Digitalität ist, dass Informationen in deskriptiver Sprache als Symbol dargestellt werden. Dieser Verarbeitungsmechanismus entspricht nicht dem des Gehirns. Das Gehirn bearbeitet Informationen analog. Das heißt physikalische Variablen wie zum Beispiel die Geschwindigkeit oder die Intensität entsprechen der eigentlichen Information. Eine Informationseinheit wird im Gehirn also nicht durch ein Symbol abgebildet, sondern die Art und Weise der Übertragung entspricht dem Symbol. Das jedoch widerspricht der Informationsverarbeitung eines Digitalcomputers, so dass DREYFUS mit der Frage endet, ob Computer überhaupt die geeigneten Maschinen sind, um die Tätigkeit des Gehirns zu simulieren.
2. ROSENBILTH beschreibt einen zweiten Argumentationsstrang gegen die biologische Annahme. Hierbei wird der Schwerpunkt nicht auf die Art der Informationsverarbeitung gelegt, sondern auf die Wechselwirkung der einzelnen Neuronen untereinander. Im Gehirn werden einzelne elektrische Stöße immer auch an benachbarte Neuronen weitergegeben. Ein Neuron steht also nicht isoliert im Raum, sondern interagiert immer mit vielen anderen Neuronen. Die Maschine bleibt jedoch nicht-interaktiv organisiert.

### 1.2.2. Die erkenntnistheoretische Annahme

„Menschliches Denken lässt sich formal beschreiben und ist mit den klassischen naturwissenschaftlichen Methoden zugänglich.“ (Daniel/Striebel 1993: S.23)

Diese erkenntnistheoretische Annahme lässt sich in zwei Behauptungen segmentieren.

1. „Jedes nicht willkürliche Verhalten ist formalisierbar.“ (Dreyfus 1985: S.138)
2. „Mit einem Formalismus kann das entsprechende Verhalten reproduziert werden.“ (Dreyfus 1985: S.138)

Das menschliche Denken sich formal beschreiben lässt, dem stimmen auch DERYFUS und DREYFUS im ersten Kapitel ihres Buches zu. „Vom Neuling zum Experten“ (vgl. Dreyfus/Dreyfus, 1988: S. 37 – 80) nennen sie einen Aufsatz, der sich mit der Formalisierung menschlicher Lern- und damit auch Denkprozesse beschäftigt. Sie haben für den Erwerb einer neuen Fähigkeit fünf Stufen angenommen. Diese sind das Ergebnis einer Studie über den Fertigkeiten-Erwerb bei Flugzeugpiloten, bei Schachspielern, Autofahrern und Erwachsenen, die eine zweite Fremdsprache lernen.

#### 1. Stufe: Neuling

In der ersten Stufe macht man erste Versuche mit der neuen Fertigkeit umzugehen, hierzu dienen auf der einen Seite Instruktionen, auf der anderen Seite erlernt man Regeln, die kontextfrei formuliert werden, das heißt, sie sind aus der Gesamtsituation herauszuziehen und gelten immer. Zum Beispiel sagt der Fahrlehrer dem Neuling, er solle bei 20 km/h immer in den zweiten Gang schalten. Diese Regel ist immer anzuwenden.

#### 2. Stufe: Fortgeschrittener Anfänger

Der Mensch erwirbt mit der Zeit Erfahrungen in der erlernten Fertigkeit. Es werden auch schon Entscheidungen getroffen, die situativ sind, und sich von der kontextfreien Regel unterscheiden. Es findet ein Lernprozess statt, in dem der Lernende die Ergebnisse der situativen mit den kontextfreien Regeln vergleicht. In jeder Situation entscheidet sich der Lernende neu für die erfolgreichste Regel nach dem Prinzip von ‚trial and error‘. Für den Autofahrer gilt

zwar immer noch die Regel bei 20 km/h in den zweiten Gang zu schalten, doch nimmt er zusätzlich Motorgeräusche wahr, die es ihm ermöglichen auch ohne Blick auf den Tacho abzuwägen, wann es nötig ist, in den zweiten Gang zu schalten. Es fällt schwer, eine Regel auf der Basis eines erhöhten Motorgeräusches aufzustellen, die dem Fahranfänger deutlich macht, in den zweiten Gang zu schalten. Situative Regeln sind also schwer zu formulieren und ergeben sich ausschließlich durch die Erfahrung, die der Lernende in seiner neuen Fertigkeit erwirbt.

### 3. Stufe: Kompetenz

Die Zahl der kontextfreien und situativen Elemente werden ansteigen. Das heißt, die Situationen werden immer komplexer. Hierzu muss der Mensch die ihm gegenüberstehende Situation organisieren. Er setzt sich ein Ziel und formuliert danach einen geordneten Plan. Der Autofahrer, der von A nach B in möglichst kurzer Zeit fahren möchte, wird dementsprechend seinen Fahrstil konzipieren. Durch die bewusste Planung seiner Handlung steht der Handelnde in einem engen Verhältnis zu dem Ergebnis, erfühlt sich für den Ausgang der Handlung verantwortlich.

### 4. Stufe: Gewandtheit

Der gewandte Mensch ist in der Lage Ähnlichkeiten zu erkennen. Er kann eine aktuelle Situation mit einer früheren Situation vergleichen, ohne sie in ihre Einzelteile zu zerlegen. DREYFUS und DREYFUS bezeichnen dies als holistisches Erkennen von Ähnlichkeiten.

### 5. Stufe: Experte

Das Können ist Teil seiner Person geworden. Einen Großteil seiner Entscheidungen trifft er intuitiv. „Wenn keine außergewöhnlichen Schwierigkeiten auftauchen, lösen Experten weder Probleme noch treffen sie Entscheidungen; sie machen einfach das, was normalerweise funktioniert.“ (Dreyfus/Dreyfus, 1988: S. 55) Dennoch treffen Experten ihre Entscheidungen nicht unreflektiert, es steht jedoch nicht mehr das Problemlösen im Vordergrund, sondern eher eine kritische Betrachtung ihrer Intuition.

„Computer sind universelle Symbol-Manipulatoren. Daher können sie jeden Prozess simulieren, der sich exakt beschreiben lässt.“ (Dreyfus/Dreyfus, 1988: S. 82)

Der Annahme, menschliches Denken sei formal zu beschreiben, ist also zuzustimmen. Der ersten Behauptung, jedes nicht willkürliche Verhalten ist formalisierbar, lässt sich auch zustimmen, wenn mit „nicht willkürlich“ gemeint ist, einer strikten Regel folgend. Denn eine symbolverarbeitende Maschine kann nur das darstellen, was kontextunabhängig ist, objektiven, interpretationsunabhängigen Merkmalen folgt und in strikten Regeln zu formulieren ist, da Merkmale einem Symbol eindeutig zugewiesen werden müssen. Auch der zweiten Behauptung, dass entsprechendes Verhalten mit einem Formalismus zu reproduzieren ist, ist zuzustimmen, wie DREYFUS und DREYFUS an erfolgreichen Versuchen in definierten Mikrowelten zu belegen wissen (vgl. Dreyfus/Dreyfus, 1988: S.105 – 109). „Obwohl logische Maschinen die für den fortgeschrittenen Anfänger so wesentlichen situationalen Elemente nicht erkennen können, lassen sie sich dennoch darauf programmieren, ihre kontextfreien Regeln in Bezug auf Ziele zusammenzustellen – wie ein kompetenter Mensch. Man kann dem Computer eine Regel eingeben, die ihm sagt, dass eine Situation, wenn bestimmte Fakten vorliegen, auf ein bestimmtes Ziel hin organisiert werden soll.“ (Dreyfus/Dreyfus, 1988: S.97) Ähnlich gehen Expertensysteme im Bereich der medizinischen Diagnose vor. Dennoch ist die denkende Maschine nicht in der Lage, auch die vierte und fünfte Stufe menschlichen Umgangs mit Fertigkeiten zu erreichen. Das holistische Erkennen von Ähnlichkeiten ist, wie die nachfolgenden Ausführungen zeigen werden, nicht mit einem symbolverarbeitenden Digitalrechner nachzuvollziehen:

DREYFUS und DREYFUS erklären die Funktionsweise eines holographischen Mustererkenner, um deutlich zu machen, wie sich die mechanistische Vorstellung über die Arbeitsweise des Gehirns von der holistischen unterscheidet. Hologramme entstehen, wenn sich zwei Lichtstrahlen, die ein Objekt anstrahlen, überlagern, dadurch entsteht ein neues holographisches Bild zum Beispiel auf einer Fotoplatte. Überlagert man nun zwei Hologramme miteinander, so entstehen dort weiße Flecken, wo das eine Bild mit dem anderen übereinstimmt. Dieses holistische Erkennen von Ähnlichkeiten unterscheidet sich jedoch von dem einer symbolverarbeitenden Maschine, die jedes vorhandene Merkmal abglei-

chen muss. Wie die nachfolgende ontologische Annahme zeigen wird, ist die KI-Forschung sogar darauf angewiesen, jede Situation in möglichst kleine Elemente zu zerteilen, damit ihnen eindeutig ein Symbol zugewiesen werden kann. Das menschliche Gehirn scheint also eher holographisch als mechanistisch zu arbeiten, da es dem Menschen möglich ist Situationen als Ganzes zu begreifen.

### 1.2.3. Die ontologische Annahme

Die aus KI-Sicht gemachte ontologische Annahme lässt sich in drei Aussagen zusammenfassen:

1. Die Welt besteht „aus objektiven, von Menschen und untereinander unabhängigen Fakten.“ (Daniel/Striebel 1993: S.23)
2. Die Welt lässt sich in isolierbare, kontextunabhängige, kleine Elemente, zerlegen. (vgl. Daniel/Striebel 1993: S.24)
3. Alle Elemente jedoch sind wiederum logisch voneinander abhängig. (vgl. Dreyfus 1985: S.106)

Der Ansatz der Gegenargumentation ist aus der Linguistik abgeleitet. CHOMSKY und andere Vertreter der Transformationsgrammatik haben „von menschlicher Sprachverwendung abstrahiert und so die menschliche Fähigkeit formalisieren können, grammatisch korrekte Sätze zu erkennen und inkorrekte zurückzuweisen.“ (Dreyfus 1985: S.145/146) Es bleibt also die Frage, ob der Computer in der Lage ist, das formalisierte Verhalten anschließend zu reproduzieren. Hierbei müsste es der KI-Forschung gelingen, nicht nur die Sprachkompetenz, sondern auch die Sprachverwendung zu formalisieren. In diesem Punkt jedoch scheitert wiederum die Verwirklichung regelgeleiteter Systeme, denn ein Programm wird nie den Sinn einer Aussage „in einen Kontext menschlichen Lebens“ (Dreyfus 1985: S.147) einordnen können. Aber in vielen Fällen macht erst der jeweilige Kontext eine Aussage verständlich. „Für eine vollständige Theorie der praktischen Fähigkeiten von Sprechern braucht man nicht nur grammatische und semantische Regeln, sondern darüber hinaus Regeln, die es einer Person oder einer Maschine ermöglichen würden, den Kontext zu erkennen, in dem die Regeln angewendet werden müssen.“ (Dreyfus 1985: S.151) Es müssten also Regeln für die Regeln entwickelt werden, um einen Kontext als eine Ausnahme von der Regel zu beschreiben. In diesem Moment befindet sich

das Programm in einem unendlichen Regress. „Da wir [Menschen] in der Lage sind, von unserer Sprache Gebrauch zu machen, kann dieser Regress für Menschen kein Problem sein. Wenn künstliche Intelligenz möglich sein soll, darf dies auch für Maschinen kein Problem sein.“ (Dreyfus 1985: S.151) Das sich hier an die Formalisierbarkeit von menschlichem Verhalten anschließende Problem wird von DREYFUS und DREYFUS auch das „Relevanzproblem“ (Dreyfus/Dreyfus, 1988: S.114) genannt. So ist es möglich jede in sich abgeschlossene Situation zu formalisieren. Versuche, diese vielen abgeschlossenen Situationen zu einem Gesamtverhalten zusammenzuschließen, gelangen jedoch nie. Hierzu schreibt MINSKY, dass „die in Semantic Information of Processing beschriebenen Programme am Besten arbeiten werden, wenn man ihnen die exakt notwendigen Fakten eingibt“, aber „sie werden unerbittlich stecken bleiben, wenn die Informationsdateien wachsen.“ (Minsky, 1968: S.18) Er schreibt weiter: „Jedes Programm arbeitet nur in seinem begrenzten Spezialgebiet und es gab keine Möglichkeit, zwei verschiedene Problem-Löser miteinander zu verbinden.“ (Minsky, 1968: S.13) Der Computer scheint mit wachsender Komplexität der Situation nicht mehr unterscheiden zu können, welche Regeln in einem bestimmten Zusammenhang von Bedeutung sind und welche nicht. Für einen Digitalcomputer ohne eine Beziehung zu der erlebten Welt sind die aus einem „Kontext herausgelösten Tatsachen eine sperrige Masse neutraler Daten.“ (Dreyfus 1985: S.234) So ist dem ontologischen Argument nur solange zuzustimmen, wie die programmierte Situation klar definiert ist. Es ist jedoch im Moment nicht vorstellbar, dass eine komplexe Situationen so formalisiert werden kann, dass die Maschine zwischen relevanten und irrelevanten Regeln unterscheiden kann. Grundsätzlich bleibt die Frage, ob Menschen in der Lage sind auch komplexe Situationen so zu beschreiben, dass sie komplett in Regeln gefasst werden können. „Allein aber die Anzahl amtierender Juristen zeigt uns, dass es unmöglich ist, Ambiguitäten, Ermessens- und Urteilsspielräume auszuräumen, indem man die Gesetzbücher so komplettiert, dass sie alle möglichen Situationen beschreiben und vorwegnehmen.“ (Dreyfus/Dreyfus, 1988: S.114)

#### **1.2.4. Zusammenfassende Übersicht über die Grenzen der KI**

Annahmen der KI	Aussagen der KI-Forschung über menschliche Kognitionsvorgänge	Widerlegung/Zustimmung nach DREYFUS und DREYFUS zur Übertragung auf die Fähigkeiten eines Computers
Biologische Annahme	„Das menschliche Gehirn funktioniert physiologisch wie Computerhardware.“ (Daniel/Striebel 1993: S.23)	Kein Vergleich möglich, da das Gehirn Symbole physikalisch verarbeitet (Intensität/ Geschwindigkeit), während der Computer Symbole als solche verarbeitet.
		Das menschliche Gehirn arbeitet holographisch und ist in der Lage eine Situation als Ganzes zu erkennen, der Computer arbeitet mechanistisch und ist dadurch nur in der Lage Segmente einer Situation zu verstehen.
Erkenntnistheoretische Annahme	„Menschliches Denken lässt sich formal beschreiben und ist mit den klassischen naturwissenschaftlichen Methoden zugänglich.“ (Daniel/Striebel 1993: S.23)	Dieser Aussage ist zuzustimmen: „Computer sind universelle Symbol-Manipulatoren. Daher können sie jeden Prozess simulieren, der sich exakt beschreiben lässt.“ (Dreyfus/Dreyfus, 1988: S. 82)
	„Jedes nicht willkürliche Verhalten ist formalisierbar.“ (Dreyfus 1985: S.138)	Wenn mit „nicht willkürlich“ einer strikten Regel folgend gemeint ist, ist auch dieser Aussage zuzustimmen.
	„Mit einem Formalismus kann das entsprechende Verhalten reproduziert werden.“ (Dreyfus 1985: S.138)	Auch für den Computer lassen sich kontextfreie Regeln formulieren, wobei es schon immer eine der Stärken des Computers war, Abläufe reproduzieren zu können.
Ontologische Annahme	Die Welt besteht „aus objektiven, von Menschen und untereinander unabhängigen Fakten.“ (Daniel/Striebel 1993: S.23)	Eine Aussage muss jedoch immer unter Berücksichtigung des gesamten Kontextes gesehen werden und genau da scheitern die Maschinen. Sie können vielleicht einzelne Informationen verstehen, jedoch nicht den Zusammenhang herstellen.
	Die Welt lässt sich in isolierbare, kontextunabhängige, kleine Elemente, zerlegen. (vgl. Daniel/Striebel 1993: S.24)	Der Aussage ist zuzustimmen, es ist jedoch nach DREYFUS und DREYFUS nie gelungen, diese Informationen zu einem sinnvollen Ganzen zu verknüpfen.

---

Anhand dieser abschließenden Übersicht ist zu erkennen, dass es nur wenige Grenzen gibt, die die KI zu überwinden hat. Vielfach sind diese Grenzen dabei nur dem technologischen Unvermögen zuzuschreiben. Es wird also weiterhin die Frage offen bleiben, ob es möglich ist, selbstreferentielle Systeme zu erschaffen. Hinzu kommt, dass der Ansatz der symbolischen KI, wie schon oben erwähnt, die ersten Versuche in der KI-Forschung repräsentiert. Es bleibt also abzuwarten, ob neuere Entwicklungen die von DREYFUS und DREYFUS beschriebenen Grenzen überwinden können. Dabei könnten die im Folgenden beschriebenen Expertensysteme einen Anfang einleiten.

### 1.3. Möglichkeiten der Künstlichen Intelligenz

Der Computer war bis dato immer auf einen Input angewiesen, selbst die Lösung eines Problems wurde dem Computer immer vorgegeben, er war immer eine Reproduktionsmaschine, jedoch nie eine Produktionsmaschine, so wie der Mensch. Diese und viele andere Grenzen wurden in dem vorangegangenen Kapitel erläutert, dennoch muss es einige Erfolge auf dem Gebiet der KI gegeben haben, die die mittlerweile schon 40-jährige Tradition rechtfertigen. Das nun folgende Kapitel soll einen Überblick über die Erfolge in der KI-Forschung geben, um im Anschluss daran unter Berücksichtigung der Definitionen kritisch zu analysieren, welche Fähigkeiten KI-Systeme erlangen können und in welchen Bereichen sie den Menschen nie imitieren können werden.

#### 1.3.1. Expertensysteme

Ein Expertensystem ist ein Programm, das mit Hilfe von Wenn-Dann-Regeln in der Lage ist Aussagen über einen bestimmten Sachverhalt zu treffen. Solch eine Form der Datenverarbeitung wird als wissensbasierte Verarbeitung bezeichnet. Expertensysteme sind jedoch nicht mit Datenbanken zu vergleichen, da reine Datenbanken nicht in der Lage sind ihre Daten zu interpretieren. Wissen ist also nichts anderes als interpretierte Daten. Ein wissensbasiertes System soll somit einen menschlichen Experten simulieren. „Im Unterschied zum menschlichen Experten ist das Wissen eines Expertensystems auf eine spezialisierte Informationsbasis beschränkt...“ (Mainzer, 1997: S.110)

Nach RADERMACHER ist „ein Großteil von dem, was wir tun, nicht mehr ... als Regelverarbeitung“ (Ahrweiler 1995: S.25). Hierzu werden einfache „Wenn-Dann-Regeln“ aufgestellt, die das Programm Schritt für Schritt durchläuft. Für RADERMACHER liegt der größte Beitrag der KI in den Inferenzsystemen, die geschrieben wurden, um „die Verwaltung und Abarbeitung großer Regelmengen zu erleichtern“ (Ahrweiler 1995: S.25). Schon an dieser Stelle wird deutlich, dass es, wenn auch in einem begrenzten Rahmen, möglich ist, menschliche Denkvorgänge auf eine Maschine zu übertragen.

Nachdem im Expertensystembereich sehr unterschiedliche Definitionen von Expertensystemen im Umlauf sind, erscheint es sinnvoll die differenzierende

Definition aus COY und BONSIEPEN zu übernehmen. Diese Definition trennt zwei unabhängige Aspekte, die für den Begriff Expertensystem wichtig sind, in zwei verschiedene Definitionen:

Anwendungsorientierte Definition:

Ein Expertensystem ist ein Computerprogramm, das für ein spezifisches und genau abgegrenztes Teilgebiet gewisse Aufgaben eines Experten lösen kann bzw. ihn bei der Problemlösung unterstützt. Dazu wird das spezielle Wissen des Experten explizit in symbolischer Form in einem Programm oder als Datenmenge bzw. innerhalb einer Datenbank dargestellt.

Technische Definition:

Expertensysteme sind Programme, die sich durch die Trennung der anwendungsspezifischen Methoden in der Wissensbank und der anwendungsunabhängigen Programmsteuerung durch die Inferenzmaschine zum Erzeugen logischer Schlussfolgerungen auszeichnet.

Es ergeben sich natürlich auch zwei Fälle, wo diese beiden Definitionen nicht zusammenfallen. Zum einen kann ein System Aufgaben eines Experten übernehmen und doch auf traditionelle Art programmiert sein, zum Beispiel aus Effizienzgründen. Zum anderen kann es vorkommen, dass ein System in der Art der Expertensysteme programmiert ist, dass es jedoch für Aufgaben verwendet wird, für die bisher kein Experte eingesetzt wurde, wie z.B. in der Prozesssteuerung oder in der Mustererkennung.

### **1.3.2. Funktionsweise von Expertensystemen**

Typisch für die Funktionsweise von Expertensystemen ist auf der einen Seite das Aufstellen sehr komplizierter Wenn-Dann-Regeln und auf der anderen Seite das Berechnen von Wahrscheinlichkeiten. Dadurch, dass es dynamische Regelsysteme sind, werden am Ende der Problemlösungssequenz mehrere richtige Lösungen benannt, die mit unterschiedlichen Wahrscheinlichkeiten belegt sind. Ähnlich einer quadratischen Gleichung, die immer zwei Lösungen hervorbringt,

wäre in diesem Fall die Wahrscheinlichkeit der richtigen Lösung mit 1:1 zu bestimmen. (vgl. Puppe, 1988)

Expertensysteme sind Programmstrukturen, die aus mehreren Teilen bestehen. Grundsätzlich kann man Expertensysteme in zwei „Hauptmodule“ einteilen (siehe Abb.1). Zum einen das Steuersystem und zum anderen die Wissensbasis. Zuerst soll das Steuersystem näher erläutert werden, um schließlich den Aufbau der Wissensbasis aufzuzeigen.

Das **Steuersystem** eines Expertensystems kann die Lösungsstrategie eines Experten simulieren. Die Regeln werden zu einer Konstanten, die auf unterschiedliche Wissensgebiete angewendet werden kann. Sie ist unabhängig von der Wissensbasis. Nach PUPPE besteht das Steuersystem aus 4 Komponenten (siehe Abb.1).

1. Die **Problemlösungskomponente** ist der Vermittler zwischen der Wissensbasis und dem Steuersystem. Hier wird „das Expertenwissen zur Lösung des vom Benutzer spezifizierten Problems interpretiert.“ (Puppe, 1988: S.12)
2. Die **Interviewerkomponente** liest die variablen Daten ein, bzw. die Aufgabenstellung von Seiten des Benutzers.
3. „Die **Erklärungskomponente** erläutert dem Anwender die ermittelte Lösungsstrategie.“ (Nebendahl, 1987: S. 33) So kann der Anwender Fehler in der Wissensbasis lokalisieren oder auch nur den Lösungsweg ablesen. In jedem Fall wird die Arbeit des Expertensystems transparent gemacht.
4. Die **Wissenserwerbskomponente** ist der Ort im Programm, in den der Experte sein Wissen eingeben kann.

Sind nur Wissenserwerbskomponente, Interviewerkomponente, Erklärungskomponente und Problemlösungskomponente vorhanden, spricht man von einer (Expertensystem-) Shell. Die Anwender können hier das Wissensgebiet selbst festlegen, indem sie die Wissensbank mit dem nötigen Wissen auffüllen.

„Die **Wissensbasis** eines Expertensystems enthält das Faktenwissen und Erfahrungswissen von Experten eines Aufgabengebietes.“ (Nebendahl, 1987: S. 33)

Die Wissensbasis besteht aus drei verschiedenen Teilen (siehe Abb.1):

1. Dem fallspezifischen Wissen von den jeweiligen Benutzern,
2. den Zwischenergebnissen und Problemlösungen, die von der Problemlösungskomponente hergeleitet wurden und
3. dem bereichsbezogenen Expertenwissen. (Puppe, 1988: S.12)

Während die Wissensbasis also dem Input entspricht, ist das Steuersystem das Herz des Expertensystems. Hier wird das Input so verarbeitet, wie die Regeln es vorschreiben. Erst das Steuersystem macht aus einem Expertensystem eine intelligente Maschine. Der Hauptvorteil der Trennung des Systems in Inferenzkomponente und Wissensbank ist der einer leichten Wartung und Änderbarkeit, da diese Systeme typischerweise eben in Gebieten mit sich änderndem Wissen eingesetzt werden.

Beim Erstellen eines Expertensystems vier Teilgebiete zu beachten: (vgl.Puppe, 1988: S.113ff.)

- a) Wissenserwerb
- b) Wissensrepräsentation
- c) Inferenzmechanismus
- d) Benutzerschnittstelle

zu a) Wissenserwerb

Der Wissenserwerb (vgl. Puppe, 1988: S.115ff.) ist der Versuch, das Wissen eines Experten in einer implementationsunabhängigen aber formalen Weise niederzulegen. Dies kann auf verschiedene Weise geschehen:

Indirekt:

Dazu muss der Wissensingenieur dem menschlichen Experten helfen, sein relevantes fachliches Wissen zu identifizieren, zu strukturieren und zu formalisieren. Andere Wissensquellen neben dem Experten können für den Wissensingenieur eigenes Fachwissen sowie Fallstudien oder Bücher sein.

Direkt:

„Der Experte formalisiert sein Wissen selbst.“ (Puppe, 1988: S.114) Dazu muss das Expertensystem eine leicht bedienbare Wissenserwerbskomponente haben.

Automatisch:

„Das Expertensystem extrahiert sein Wissen selbständig aus Falldaten oder verfügbarer Literatur.“ (Puppe, 1988: S.114) Diese Technik ist allerdings im Moment reiner Forschungsgegenstand.

Zum Wissen eines Experten können ganz verschiedene Dinge gehören, wie zum Beispiel Fakten, Überzeugungen, Methoden, Heuristiken und nicht zuletzt Problemlösungswissen (globale Strategien, Metawissen). Ein Phasenplan zum Wissenserwerb sieht folgendermaßen aus:

Zuerst wird ein Pflichtenheft mit organisatorischer Umgebung und Anforderungen an das Expertensystem erstellt. Anschließend wird der Grobentwurf und der zu betrachtende Realitätsausschnitt festgelegt. Danach wird das Wissen in einer, dem verwendeten Rechnersystem und der zur Anwendung kommenden Shell verständlichen Form aufbereitet. Die hierbei zustandekommenden Strukturen dienen dann der Formulierung von Regeln, die in einem letzten Teil des Phasenplans getestet werden. (nach Buchanan, 83: S.139)

## b) Wissensrepräsentation

Die Wissensrepräsentation hat eine natürliche und effiziente Darstellung des “Wissens” zum Ziel. Unabhängig von der nicht ganz klaren Bedeutung von natürlich in diesem Zusammenhang ist klar, dass diese Bedingungen eventuell in Konflikt miteinander treten können. Hierzu sind Kalküle entwickelt worden, die den regelhaften Ablauf des Systems steuern.

Ein Kalkül beschreibt, wie man aus Aussagen mit Hilfe von Ableitungsregeln neue Aussagen erhält. Vorgegebene Aussagen sind die Axiome (Fakten, Annahmen, das was nicht in Frage gestellt wird). Abgeleitete Aussagen sind Theoreme (Schlussfolgerung). PUPPE benennt sechs unterschiedliche Eigenschaften von Kalkülen:

1. Adäquatheit: Natürlichkeit der Beschreibung der Welt
2. Effizienz: Relevanz der Schlussfolgerungen für die Welt
3. Mächtigkeit: Repräsentierbarkeit von Aussagen über die Welt
4. Entscheidbarkeit: Ein Kalkül verfügt dann über die Eigenschaft der Entscheidbarkeit, „wenn für eine beliebige Aussage entschieden werden kann, ob sie aus den Axiomen folgt oder nicht.“ (Puppe, 1988: S.18)

5. Vollständigkeit: „Ein Kalkül ist dann vollständig, wenn alle Schlussfolgerungen, die semantisch (zur Welt gehörig) gelten, auch syntaktisch (im Kalkül befindlich) herleitbar sind.“ (Puppe, 1988: S.18)
6. Konsistenz: Die Aussagen dürfen sich nicht widersprechen.

c) Ein Inferenzmechanismus

Ein Inferenzmechanismus ist repräsentationsabhängig. Dabei heißt Inferenz allgemein, dass aus vorhandenem Wissen Neues erschlossen wird. Nebenbei angemerkt können die Schlussverfahren bzw. das Wissen auch vage und unsicher sein.

In diesem Zusammenhang ist es wichtig, den Begriff „Regel“ kurz zu erläutern. „Da Experten ihr Wissen oft in Form von Regeln formulieren, sind Regeln die verbreitetste Wissensrepräsentation in Expertensystemen.“ (Puppe, 1988: S. 21) „Eine Regel besteht aus einer Vorbedingung und einer Aktion.“ (Puppe, 1988: S. 21) PUPPE benennt zwei Arten der Aktionen. Zum einen die Implikation/Deduktion, und zum anderen Handlungen. Erstere prüfen den „Wahrheitsgehalt einer Feststellung“ (Puppe, 1988: S. 21), während die Handlungen einen Zustand verändern können.

Zur Abarbeitung der Regelmengen stellt sich die Frage, ob die Regeln vorwärts- oder rückwärtsverkettet bearbeitet werden (siehe Abb.2). Eine Vorwärtsverkettung kann Schlussfolgerungen nur mit einer vorgegebenen Datenbasis ermöglichen. Hierbei werden zuerst alle Schlüsse errechnet, die sich aus der Wissensbank zusammen mit den fallspezifischen Fakten ergeben. Bei der Vorwärtsverkettung gibt es zwei verschiedene Phasen, um die relevanten Regeln zu finden. Zuerst sucht das System während einer Vorauswahl innerhalb der gesamten Datenbasis. In einer zweiten spezifischeren Auswahl können dann mit Hilfe von verschiedenen formalisierten Konfliktlösungsstrategien Regeln gefunden werden, die das gestellte Problem lösen können. (siehe Abb.2)

Bei der Rückwärtsverkettung geht man von einer Endhypothese aus und versucht Regeln zu finden, die diese Hypothese aus den bekannten Regeln herleiten. Hierbei ist der Prozess der Problemlösung dialogisch. Nach der Zielformulierung des Benutzers überprüft das System die Datenbasis nach relevanten Regeln, wenn das Problem lösbar ist, werden mit ähnlichen Konfliktlösungsstrate-

gigen Regeln gefunden, wie bei der Vorwärtsverkettung. Ist das Problem mit der vorgegebenen Datenbasis nicht zu lösen, wird ein Unterziel formuliert und das System braucht einen erneuten Input durch den Benutzer. Diese Schleife vollzieht sich solange, bis dem System alle Daten zur Verfügung stehen, um zu einer Schlußfolgerung zu kommen. (siehe Abb.2)

Die Vorwärtsverkettung ist vorteilhaft, wenn alle Daten von Anfang an vorhanden sind (z.B. Konstruktion) bzw. wenn auf neu ankommende Daten reagiert werden muss (z.B. Prozessüberwachung).

Die Rückwärtsverkettung hat Vorteile, wenn nur eine kleine Zahl von Endhypothesen vorhanden ist wie z.B. bei manchen Diagnose- und Klassifikationsaufgaben. Diese Methoden können bei Bedarf auch kombiniert werden. (vgl. Puppe, 1988: S.21ff.)

#### d) Die Benutzerschnittstelle

Die Benutzerschnittstelle muss mit zwei Anwendersichten konstruiert werden: Zum einen die Sicht für den Experten bzw. Wissensingenieur beim Aufbau und der Wartung der Wissensbank, zum anderen die Sicht des Nutzers in der Anwendung des Systems.

Bei Expertenschnittstellen (siehe Abb.1) sind zum Beispiel für die Eingabe von Wissen Regeleditoren üblich, oder die Implementation von formalen Sprachen zur Beschreibung von Regeln und Fakten. Als Forschungsgegenstand gibt es Versuche natürlichsprachlich - z.B. aus Texten - Wissen in das System zu übertragen. Ein anderer Forschungsgegenstand sind sogenannte lernende Systeme, die zumeist Regeln aus Beispielen selbständig extrahieren können sollen.

Ein wichtiger Aspekt des Wissenserwerbs ist die Sinnfälligkeitsprüfung, da neues Wissen mit dem alten in Widerspruch treten kann und immer wieder Seiteneffekte zum Beispiel durch neue, geänderte oder entfernte Regeln auftreten können. Es gibt noch wenig Methoden, die diese Problematik wirklich lösen können.

Die Schnittstelle Benutzersystem (siehe Abb.1) ist eine Dialogkomponente zur Problemformulierung für Rückfragen des Systems, für Fragen des Benutzers über den Lösungsweg und schließlich zur Ergebnisausgabe.

Angestrebt wird dabei von vielen KI-Forschern die möglichst durchgängige Verwendung von natürlicher Sprache. Beim jetzigen Stand der Technik ist dies jedoch reiner Forschungsgegenstand.

Die Ergebnisdarstellung kann dabei auch grafisch sein. Wichtig ist in einigen Fällen eventuell die Unterdrückung von Details, also eine Art Ergebnisabstraktion.

Die Erklärungskomponente liefert als Antwort zumeist eine Art Protokoll der Inferenzschritte. Dieses Protokoll wird oft noch aus Gründen der Verständlichkeit weiter aufbereitet, da das einfache Mitschreiben der Inferenzschritte für den Benutzer nicht oder nur sehr schwer verständlich ist. Diese Komponente sollte optimalerweise immer zugänglich sein. (siehe Abb.1)

Man kann der Erklärungskomponente zumeist zwei Typen von Fragen stellen: Wie-Fragen (Wie kommt es zu diesem abgeleiteten Faktum?) und Warum-Fragen (Warum wird diese Zwischenfrage gestellt?). (vgl. Puppe 1988: S.132ff.)

### **1.3.3. Anwendung von Expertensystemen**

Expertensysteme sollen dort eingesetzt werden, wo man auf regelgeleitete, komplexe Systeme stößt, die von Menschen unwirtschaftlich und oft fehlerhaft bewältigt werden. Im Umgang mit Expertensystemen sind drei Personenrollen vorgesehen: Benutzer, Experte und Wissensingenieur. Verschiedene Rollen können aber auch in einer Person zusammenfallen. Zum Beispiel wenn der Experte selbst sein Wissen formalisiert und in das System eingibt.

Expertensysteme können für viele verschiedene Bereiche erstellt werden. Sie sind aber vor allem gute Problemlöser für wohldefinierte Aufgabengebiete. Folgende spezifische Kategorien haben sich herausgebildet (vgl. Busch/u.a., 1994: 134ff.; Savory, 1985; Herzog/u.a., 1993: S.125ff.; Häuslein, 1993: S.61ff.):

#### **1. Analyse- und Interpretationssysteme:**

Große Informationsmengen werden geordnet, analysiert, reduziert und nach Bedarf aufbereitet.

Beispiel: WWW Search Engines, Recherchen in Online-Datenbanken,...

## 2. Simulationssysteme:

Simulationen sind modellhafte Abbildungen eines realen Phänomens (vgl. Kap. 2.3.2.2.) HÄUSLEIN definiert ein Simulationssystem folgendermaßen: „Ein Simulationssystem ist ein Softwaresystem, das die rechnergestützte Bearbeitung der drei Aufgabenbereiche Modellbildung, Durchführung von Simulationsexperimenten und Ergebnisanalyse im Rahmen einer Simulationsstudie unterstützt.“ (Häuslein, 1993: S.68)

Beispiel: Jäger-Beute-Simulationen in der Biologie, Simulationssysteme im Umweltbereich z.B. zur Erstellung eines Wetterberichtes

## 3. Diagnose- und Fehlersuchsysteme:

Daten über Zustand und Verhalten des zu diagnostizierenden Objekts werden vom System erfragt. Auf dieser Informationsbasis zieht das System Schlüsse über eine etwaige Fehlfunktion. Diese Art von wissensbasierten Systemen werden häufig auch von Experten selbst genutzt. (vgl. Busch/ u.a., 1994: S.134ff.)

Beispiel: medizinische Diagnose, Risikoprüfung in der privaten Krankenversicherung

## 4. Beobachtungs- und Kontrollsysteme:

Das Expertensystem dient als Feedback-Mechanismus, der die über Sensoren erhaltenen Informationen auswertet und danach Prozesse mit Kontrollsignalen steuert.

Beispiel: Steuerung und Überwachung von chemischen Prozessen

## 5. Designsysteme:

Das Expertensystem erhält Spezifikationen von Produkten oder Bauteilen. Es konstruiert selbständig und zeichnet unter Zuhilfenahme von CAD-Software.

Beispiel: Schaltungsentwurf und -zeichnung

## 6. Wirtschaftswissenschaftliche Systeme

Viele wirtschaftswissenschaftlichen Abläufe lassen sich in Regeln formulieren. HERZOG vermutet aus diesem Grund, „dass innerhalb der Informationsbear-

beitung die wissensbasierten Techniken einen interessanten Beitrag zur Kundenbedienung leisten können.“ (Herzog/u.a., 1993: S.127)

Beispiel: Berechnung des Beitrags für Autoversicherungen, Erstellung von Personalplänen

#### 7. Tutorielle Systeme:

Das Expertensystem dient als Lehrer, der sich mit einer bestimmten Wissensdomäne vertraut macht.

Beispiel: SOPHIE lehrt Technikern die Fehlersuche in elektronischen Schaltungen

### **1.3.4. Bewertung der KI in Bezug auf die Implementierung in ein Lernprogramm**

Wie gezeigt wurde, sind Programme, die mit künstlicher Intelligenz ausgestattet sind, in der Lage, alle die Tätigkeiten des Menschen zu imitieren, die auf Regeln basieren. Sie sind jedoch nicht in der Lage, wie auch SEARLE schon feststellte, zu verstehen. Er geht dabei auf das Imitationsspiel von Turing ein. Seiner Meinung nach wird es in Zukunft mit Sicherheit Maschinen geben, die in der Lage sein werden Denkaufgaben und Vorgänge in der Weise zu erledigen, wie sie heute unser Gehirn erledigt. Computer zählen für ihn aber nicht zu diesen Maschinen. Um seine These zu stützen, bringt er das Beispiel mit dem chinesischen Zimmer. Hierbei gibt es einen Menschen, dessen Muttersprache Englisch ist. Dieser sitzt in einem geschlossenen Raum. Vor einem Fenster außerhalb des Raums stehen Personen deren Muttersprache chinesisch ist. Der Mensch in dem Raum erhält nun mehrere Symbole, die die Außenstehenden als chinesische Schriftzeichen identifizieren würden. Als Nächstes erhält die Person im Zimmer eine Anleitung auf Englisch, seiner Muttersprache, wie er die Symbole zu kombinieren hat. Außerhalb des Raumes erkennen die Personen Sätze, bzw. Aussagen. Und als Letztes erhält die Person im Raum noch eine Anleitung in seiner Muttersprache Englisch, wann er welche Schriftzeichenkombinationen zu verwenden hat. Die Personen außerhalb spielen mit der Person im Raum nun das Imitationsspiel. Aufgrund der Anleitungen, wann welche Symbole zu verwenden sind, ist die Person im Raum nun in der Lage das Imitationsspiel problemlos zu meistern. Stellt man jedoch die Frage, ob er etwas ver-

standen hat, von dem was er mit den Symbolen zum Ausdruck brachte, antwortet dieser mit nein. Er hat lediglich die Anweisungen in der Sprache verfolgt, die er seine Muttersprache nennt. Für die Außenstehenden ist diese Tatsache nicht erkennbar. Für sie hat er zu jeder Frage die passende Antwort gegeben. SEARLE vergleicht nun die Person im Raum mit einem Computer (der Hardware) und die Anweisungen mit dem Programm (der Software). So steht für ihn fest, dass Computer, wie wir sie heute kennen, nicht in der Lage sind zu verstehen. Diese Ansicht wird auch starke KI genannt. Den Aussagen der schwachen KI, Computer wären in der Lage gewisse Denkprozesse zu simulieren, stimmt SEARLE jedoch zu, womit wir bei den Möglichkeiten von KI angelangt sind.

Nach der Recherche der unterschiedlichen Definitionen von Intelligenz, ist aufgefallen, dass Denkprozesse oft dann als intelligent bezeichnet wurden, wenn sie so komplex sind, dass die Kognitions-Forscher sie nicht mehr erklären können (Bewusstsein, Intention). Sobald jedoch ein neues Erklärungsmodell gefunden wird, lüftet sich der Zauber der Intelligenz und es bleibt eine nackte Theorie zurück, nach der man den erklärten Denkprozess nicht mehr als intelligent bezeichnen würde. Ist man jedoch erst in der Lage, das erkannte Phänomen in einer Maschine nachzubauen, bleibt nichts weiter zurück als ein Formalismus. Es wird folglich immer schwerfallen, eine Maschine als intelligent zu bezeichnen, weil sie immer den Menschen abbilden wird. (vgl. Schanda, 1995: S.62) Eine Maschine, die dem menschlichen Gehirn nachempfunden ist, wird es erst dann geben, wenn man die Arbeit des Gehirns vollständig erforscht hat. Deshalb behelfen sich die Forschungsbereiche psychologische Kognition, Neurophysiologie und KI gegenseitig (Bibel/u.a., 1987: S.1ff.), um auf der einen Seite das Geheimnis der Intelligenz zu lüften und auf der anderen Seite „intelligente“ Maschinen zu bauen.

Wenn man die KI in ein Lernprogramm implementieren will, wird man also auf die erfolgreichen Versuche der KI-Forschung zurückgreifen müssen, auf die Expertensysteme. Die Frage, die sich stellt, ist, ob Expertensysteme so zu konstruieren sind, dass sie menschliches Lernen unterstützen können. Dazu müsste sich das System dem Lerntypus, dem Lernniveau und dem Vorwissen des Lernenden anpassen können. Wie oben schon erwähnt, sind zumindest die rückwärtsverketteten Regelsysteme in der Lage, einen Dialog mit dem Benutzer

auszuführen, und können die neu gewonnenen Erkenntnisse in die Regelbasis integrieren. PUPPE schreibt dazu: „Da das Wissen in guten Expertensystemen strukturiert aufbereitet ist und auch viele praktisch nützliche Heuristiken enthält, ist der Einsatz für tutorielle Zwecke naheliegend. Dabei kommen jedoch neue Anforderungen auf das Expertensystem zu.“ (Puppe, 1988: S.137) Die neuen Anforderungen, die PUPPE benennt, sollen in einem kurzen Abriss wiedergegeben werden (Puppe, 1988: S.137ff.):

1. Es sollte eine Ähnlichkeit in der Problemlösungsstrategie zwischen Lernendem und dem Expertensystem bestehen, damit die Aussagen der Erklärungskomponente den Benutzer nicht verwirren, sondern den Fehler des Lernenden aufgreifen und erklären können.
2. Die Wissensbasis muss für den Benutzer individuell zu erarbeiten sein (Hypertext).
3. Die Problemlösung sollte vom Benutzer ausgehen, während das System zwar parallel das Problem verarbeitet, um dann jedoch dem Lernenden ein Feedback geben zu können. Das heißt, das System muss über eine Vergleichskomponente verfügen, mit dessen Hilfe die Problemlösung des Benutzers mit der des Expertensystems abgeglichen werden kann.
4. Damit mehrere Personen das System benutzen können, sollte das Programm in der Lage sein, benutzerspezifische Daten abzuspeichern, da der Lernprozess ein individueller Ablauf ist, an dessen Ende zwar meist ein ähnliches Ergebnis steht, dessen Verlauf jedoch sehr unterschiedlich ist.

Solche Lernprogramme wurden unter dem Namen IT-Systeme (intelligente tutorielle Systeme) entwickelt. Welche Eigenschaften ein solches Lernprogramm hat, wird im nächsten Kapitel beschrieben. Hier soll die Eigenart der menschlichen Lernprozesse dargestellt werden, und die sich daraus ergebenden Konsequenzen für die Modellierung eines Lernprogramms auf lerntheoretischer Basis.