

Stefan Meretz (Januar 1997)

## **KI? KI! Von der "Künstlichen Intelligenz" zur *Kritischen Informatik***

**Die Forschungsrichtung der "Künstlichen Intelligenz" (KI) ist umstritten, weil sie uns "nahe geht". Die KI verwendet keine grundsätzlich anderen Techniken als die übrige Informatik. Dennoch meinen Kritikerinnen und Kritiker, der KI müsse Grenzen gesetzt werden. Sie unterscheiden damit unausgesprochen zwischen "normaler Informatik" und KI, zwischen mittelbar auf Mensch und Gesellschaft wirkender Technik und unmittelbar mit dem Menschen befaßter Technik.**

Kritikerinnen und Kritiker der KI melden vorwiegend moralische Bedenken an, Begriffe wie "Verantwortung" und "Ethik" stehen dabei im Mittelpunkt. Eine solche Kritik muß jedoch schwach bleiben und ein unbefriedigendes Gefühl hinterlassen. Die Empörung ist groß, wenn die Moralkriterien von KI-Fans nicht angenommen werden. Natürlich fand der KI-Kritiker Weizenbaum das Buch "Mind Children" von Moravec [1] "grauenhaft", da Moravec das vertritt, wovor Weizenbaum den Zeigefinger hebt [2]. Die Abgeklärtheit, zu sagen "Lustiges Buch, ich liebe Science Fiction" kann Weizenbaum nicht aufbringen - aber vielleicht liebt er ja wirklich nicht Science Fiction.

Moral ist Ausdruck gesellschaftlicher Praxis - und nicht umgekehrt. Eine Moral zu definieren, um eine moralgeleitete Praxis zu bewirken, funktioniert fast nie. Der Moralismus der KI-Kritikerinnen und -Kritiker ist Ausdruck ihres Unbehagens und als solches auch ernstzunehmen - Grundlage einer wirkungsvollen Kritik kann er nicht sein. Der an die Wand gemalte Teufel lacht die Kritikerinnen und Kritiker aus - in früheren Zeiten sollte die Eisenbahn das Ende der Zivilisation einläuten, da der Mensch nicht für so große Geschwindigkeiten geschaffen sei. Die Verwendung von Moral als Hauptinstanz der Kritik bedeutet auch das Eingeständnis, sachlich der KI-Denke letztlich nicht beikommen zu können. Haben die KI-Fans also eigentlich Recht?

Nein, dem ist nicht so. Eine sachliche, immanente Kritik ist machbar, doch: Wer von der KI reden will, darf von der Informatik nicht schweigen. Die KI ist kein exterritoriales Gebiet, sondern Bestandteil der Informatik. Sie gründet in ihrem Paradigma und verwendet ihre Theorien. Was die KI interessant macht, ist die Zuspitzung einiger Grundannahmen der Informatik. Dadurch werden einige Aspekte "sichtbarer". Es gilt jedoch auch festzuhalten: Informatik ist *nicht gleich* KI.

Im folgenden soll *ein* Kernkonzept der Informatik und KI diskutiert werden: das der *Bedeutungen*. Bedeutungen - oft wird gleichbedeutend der Begriff "Informationen" verwendet - sind für die Informatik und die KI ein zentrales Problem. Moravec will das Bedeutungsproblem seiner KI-Roboter lösen, indem er "Wahrnehmungselemente" in ein "Verarbeitungsmodell" umformt, das wiederum mit "Interaktionselementen" verknüpft ist. Auf diese Weise könnten die Roboter "alle Gegenstände als das erkennen, was sie sind" [3]. In der Softwareentwicklung soll die bedeutungsvolle Welt in einem Anwendungsbereich analysiert und dann mit Hilfe von CASE-Tools in ein formales Modell transformiert werden, auf dessen Grundlage schließlich die Software entsteht. In beiden Fällen geht es um den gleichen Vorgang, um die Zuordnung von Bedeutungsvollem zu Formalem, von Bedeutungen zu Zeichen. Dem liegt die Annahme zugrunde, das man die Bedeutungsentstehung real genau durch einen solchen Zuordnungsakt erklären kann. Diese *Zuordnungshypothese* wird in seltener Klarheit von Kayser [4] vertreten (siehe Kasten "Zuordnungshypothese").

Die Zuordnungshypothese vermischt zwei Bedeutungs-Welten - die unabgeschlossene, bedeutungsvolle Welt des Menschen und die Syntax-Semantik-Welt des Programms. Für die menschliche Welt wird angenommen, daß sich die Menschen auf die Bedeutungen "einigen". Eine so

verstandene "Pragmatik" ist jedoch zirkulär und trivial: Woran erkenne ich einen Stuhl, um ihn als Stuhl zu bezeichnen? Die Frage setzt das Resultat voraus, die zu erklärende Bedeutung ist in der Frage als symbolischer, "gedachter Stuhl" bereits gegeben. Und die Antwort ist trivial: Ein Stuhl ist ein Stuhl. Aus diesem Zirkel gibt es auch dann kein Entkommen, wenn man viele Menschen einbezieht oder den Stuhl in Teile "zerlegt".

Die informatische Deutungsweise der Zuordnungshypothese ist weit sinnvoller. Dort ist das Syntax-Semantik-Verhältnis im Unterschied zum menschlichen Welt-Raum auf einen eindeutigen *Konventionsraum* beschränkt (siehe Kasten "Konventionsräume"). Innerhalb des Konventionsraums sind die Syntax-Semantik-Relationen fest definiert (worden). Das ist vergleichbar mit physikalischen oder chemischen Prozessen, denen das Ursache-Wirkungsschema zugrunde liegt. Doch so wie der Mensch nicht als Ursache-Wirkungs-Mechanismus erklärbar ist, so ist die bedeutungsvolle Welt des Menschen nicht als Syntax-Semantik-Schema abbildbar. Genau das aber versucht die KI! Sie verwendet die Begriffe "Informationsverarbeitung" und "Symbolmanipulation" für den Konventionsraum und die Lebenswelt des Menschen und verwischt damit jeden Unterschied. "Informationsverarbeitung" erscheint nurmehr als ein Problem der "Quantität". So kann Moravec zu der naiven Annahme gelangen, daß die Erzeugung von künstlicher Intelligenz von der Verarbeitungsgeschwindigkeit der Computer abhängt [3].

### **Kritik ist machbar**

Die KI wird kritisierbar, wenn es gelingt, ihr sachlich Grundannahmen zu entziehen. Am Beispiel der Bedeutungen, wurde gezeigt, daß KI und Informatik ihren Bedeutungsbegriff aus dem Syntax-Semantik-Verhältnis bei Programmen verwenden und unzulässigerweise auf menschlich-weltliche Bedeutungen ausdehnen. Programmierte Syntax-Semantik-Relationen und menschliche Bedeutungswelt sind jedoch - mit einer Computermetapher gesprochen - völlig "inkompatibel".

Notwendig ist eine Theorie der Informatik - bisher gibt es sie nicht. Eine solche Theorie kann nur eine *kritische Informatik* sein. Sie muß sich mit vorhandenen Ansätzen auseinandersetzen, ihre Grundannahmen "freilegen", ggf. sachlich begründet verwerfen und konzeptionelle Alternativen schaffen. Die KI ist dabei nicht mehr als ein bunter Vogel. Praxisrelevant wird es, wenn es um die Softwareentwicklung und andere Kerndisziplinen geht.

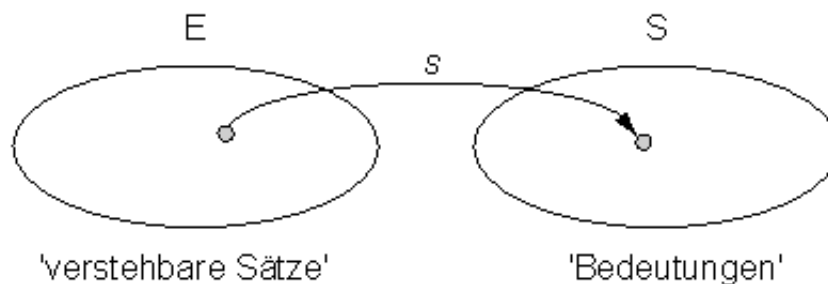
Kasten 1:

### Zuordnungshypothese

Die Zuordnungshypothese ist der Kern der informatischen Auffassung der Definition von "Bedeutung". Besonders klar hat diese Hypothese Kayser [4] formuliert:

"Die folgende These (T) wird allgemein als zutreffend angesehen: (T) Jeder Satz, der verstehbar ist, hat eine (oder, im Fall von Mehrdeutigkeit, mehrere) Bedeutung(en), und es existiert ein Formalismus - unterschiedlich zum Satz selbst -, der die Fähigkeit besitzt, sie zu repräsentieren. In mehr technischen Ausdrücken lautet die gleiche These:

Es existiert ein Raum  $S$ , der Bedeutungsraum genannt wird, und eine Abbildung  $s$ , die für jedes Element der Menge  $E$  verstehbarer Sätze ein oder mehrere Elemente von  $S$  liefert."



Die Zuordnungshypothese als Diagramm (Abb. 1).

Die Zuordnungshypothese wird in Abb. 1 grafisch veranschaulicht. Der Aufbau der These ist vollständig zirkulär. Sie "erklärt" sich durch sich selbst, in dem sie einen eigenen, abgeschlossenen Raum definiert. Die These kann in zwei Weisen - einer trivialen und einer informatischen - interpretiert werden.

#### Triviale Interpretation

Die Menge  $E$  enthält Sätze, denen mit der Abbildung  $s$  Bedeutungen - Elemente aus  $S$  - zugeordnet werden. Diese Zuordnung ist sinnvoll, wenn man annimmt, daß die Sätze durch diese Abbildung ihre Bedeutung erhalten, folglich vorher keine besitzen. Wie aber funktioniert  $s$ , wie werden die *richtigen* Sätze erkannt, denen *ihre* Bedeutung zugeordnet werden soll? Sie können erkannt werden, weil sie *verstehbar* sind. Warum sind Sätze verstehbar? Weil wir ihre Bedeutung erkennen - bedeutungslose Sätze sind nicht verstehbar. Damit ist die Bedeutung der Sätze, die erst durch die Abbildung  $s$  auf  $S$  geleistet werden sollte, durch ihre Verstehbarkeit bereits vorgegeben. Damit ist  $s$  überflüssig oder etwa so sinnvoll wie programmiertechnisch die Zuweisung  $i:=i$ . Die Mengen  $E$  und  $S$  enthalten demnach beide Bedeutungen, ihre Zuordnung ist somit trivial. Daran ändert sich nichts, wenn man annimmt, daß es sich um unterschiedliche Bedeutungstypen handelt, also etwa gegenständliche Bedeutungen (der sinnlich benutzbare "Stuhl") und symbolische Bedeutungen (das Wort "Stuhl"): Ein Stuhl ist ein Stuhl.

## **Informatische Interpretation**

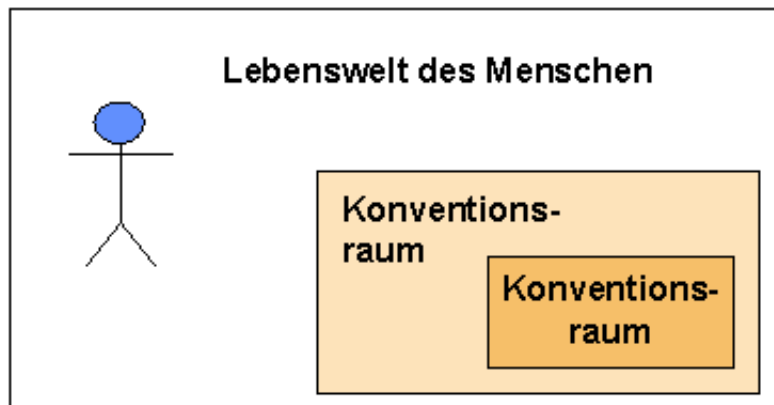
Nimmt man an, beide Mengen enthalten keine Bedeutungen im menschlich-weltlichen Sinne, sondern eine andere "Art" von "Bedeutungen". Als Beispiel mag ein Programmfragment dienen. Der "Satz" `printf("Hello world")` hat ggf. bei der Ausführung des Programms die Ausgabe des Textes "Hello world" auf dem Bildschirm (o.ä.) zur Folge. Eine Zeile eines Programms aus der Menge E "bedeutet" demnach eine Aktion aus der Menge S bei der Ausführung. Für die Zuordnung *s* sorgt der Compiler, der einen Text in Maschinencode "übersetzt", der eine Aktion bewirkt (eigentlich eine doppelte Zuordnung, aber diese Unschärfe sei gestattet). Ein anderer Compiler mit einer anderen Zuordnung *s* könnte eine andere Aktion bewirken - etwa eine Geschirrspülmaschine anschalten. In der Informatik gibt es für dieses Verhältnis von Text und Aktion das Begriffspaar *Syntax* und *Semantik*. Die Syntax legt die korrekte Form fest (oben fehlt z.B. das Semikolon!), die Semantik legt die verbundene Aktion fest. Und im Informatik-Duden [5] steht für "Semantik": "Lehre von der inhaltlichen Bedeutung einer Sprache."

## **Herkunft der Bedeutungen**

Beide Deutungsweisen der Zuordnungshypothese - die triviale und die informatische - sind Realität. Die triviale Bedeutungszuordnung vollziehen wir tagtäglich, in dem wir mit Gegenständen umgehen und über sie mit anderen sprechen, Texte lesen etc. Ohne diese "Trivialität" der permanenten Zuordnung von Bedeutungen zu Bedeutungen würde die Gesellschaft nicht funktionieren. Daraus folgt, daß "Bedeutung" *nicht* ein "platonischer" Begriff sein kann, mit dem eine "Wesensverwandtschaft" zwischen Ding und Symbol gefaßt wird [6]. "Bedeutung" muß als Vermittlungsbegriff aufgefaßt werden. Bedeutungen vermitteln zwischen den "Dingen" und uns. Bedeutung ist demnach nichts absolutes, Bedeutung haben die "Dinge" immer *für uns*. Dabei haben sich die gesellschaftlichen, also allen gemeinsamen Bedeutungen historisch herausgebildet. Sie entstanden nicht - wie oft vermutet wird - aus einer Art gesellschaftlicher "Einigung". Eine solche Annahme würde die gleiche Zirkularität enthalten, wie sie in der "trivialen" Deutungsweise der Zuordnungshypothese beschrieben wurde. Die Bedeutungen entstanden aus Notwendigkeiten bei der gesellschaftlichen Herstellung des Lebens, denn gesellschaftlich werden Dinge nicht bloß individuell oder für einen persönlich Bekannten hergestellt, sondern allgemein "für andere". Der allgemeine Andere kann das Hergestellte jedoch nur gebrauchen, wenn er die Brauchbarkeit - die Bedeutung - erkennt. Die allgemeine Brauchbarkeit entsteht in der Produktion durch sein allgemeines "Gemachtsein" für Andere. Das bedeutet, daß die (gegenständliche und mit Sprache/Schrift auch die symbolische) Bedeutung gesellschaftlich *hergestellt* wird und auf diese Weise unsere Gesellschaftlichkeit erst ermöglicht (z.B. im Unterschied zu Tieren - ausführlicher in [7]).

## Konventionsräume

In Konventionsräumen sind die Syntax-Semantik-Verhältnisse eindeutig definiert. Jedem Zeichen (Wort etc.) ist eine eindeutige Aktion beigelegt. Bei Programmen spiegeln Konventionsräume die physikalische Ursache-Wirkungs-Kausalität des Computers wider. "If-Then" ist das textuelle Gegenstück zur "Wenn-Dann-Schaltung" der Hardware des Computers. Ohne Konventionsräume keine funktionierenden Computer. Der KI-ler Haugeland hat also Recht, wenn er sagt: "Wenn man auf die Syntax achtet, wird die Semantik auf sich selbst achten?" [8]. Außerhalb von fest definierten Konventionsräumen ist diese Aussage jedoch ungültig. Die Ausdehnung der Computerlogik auf die menschliche Welt ist folglich unzulässig. Die menschliche Welt ist hinsichtlich ihrer Bedeutungen unbegrenzt und unabgeschlossen, und es besteht auch keine feste Zuordnung von Zeichen zu Bedeutungen. Das macht im übrigen auch die Begrenztheit von maschinellen Übersetzungsprogrammen aus.



Nur *innerhalb* eines Konventionsraums gilt die Abbildung: Syntax  $\leftrightarrow$  Semantik (Abb. 2)

Konventionsräume können ein hierarchisches Verhältnis aufbauen. Gleich "Schachteln" in "Schachteln" (vgl. Abb. 2) definieren sie auf jeweils höheren Ebene komplexere Syntax-Semantik-Verhältnisse. Ein Beispiel ist etwa die Schachtelung Maschinencode - Assembler - Hochsprache - CASE-Diagramm. Wenn Syntax-Semantik-Verhältnisse in Konventionsräumen eingeschlossen bleiben - was sie müssen, damit Computer funktionieren -, dann ist es unerheblich, mit welcher Geschwindigkeit die Zeichenketten in Aktionen umgesetzt werden. Es ist auch unerheblich, ob es sich bei den Aktionen um Bildschirmausgaben oder Bewegungen von Motoren von Robotern handelt. Die "höchste", komplexeste "Schachtel" kann nie in Richtung bedeutungsvoller Menschenwelt überschritten werden. Das wäre gleichbedeutend mit einer Auflösung des Konventionsraumes und damit der Bedingung, die Computer erst funktionieren lassen. Wenn man nicht von der These ausgeht -Der Mensch denkt, Gott lenkt? im unmittelbar technischen Sinne, dann ist die äußerste "Schachtel" auch die letzte - und die hat keine Wände!

## Literatur

- ↑ [1] Hans Moravec, Mind Children, Hoffmann und Campe, Hamburg 1990
- ↑ [2] Gespräch mit Hans Moravec, c't 6/96, S. 106ff.
- ↑ [3] Hans Moravec, Geisteskinder, c't 6/96, S. 98ff.
- ↑ [4] D. Kayser, A computer scientist's view of meaning, in: S.B. Torrance, The Mind and the Machine. Philosophical Aspects of Artificial Intelligence, Ellis Horwood, Chichester 1984
- ↑ [5] Duden Informatik, Dudenverlag, Mannheim/Wien/Zürich 1989
- ↑ [6] John Lyons, Einführung in die moderne Linguistik, Verlag C. H. Beck, München 1989
- ↑ [7] Anita Lenz, Stefan Meretz, Neuronale Netze und Subjektivität, Vieweg-Verlag, 1995.
- ↑ [8] John Haugeland, Künstliche Intelligenz - Programmierte Vernunft?, MacGraw-Hill, München 1987