

Der Mensch als Turing-Maschine?

Die Frage nach der künstlichen Intelligenz in philosophischer
und theologischer Perspektive

Dr. Dirk Evers, Stäudach 39, D-72074 Tübingen

Die Frage nach der Möglichkeit und den Grenzen künstlicher Intelligenz stellt die Frage nach der Simulierbarkeit geistiger Prozesse durch von Menschen hergestellte Maschinen. Ihre Beantwortung wird mit davon abhängen, wie wir die Frage nach uns selbst beantworten, umgekehrt aber auch auf unser Selbstverständnis zurückwirken. In diesem Aufsatz soll sie so behandelt werden, dass in einem ersten Teil nach den Möglichkeiten künstlicher Intelligenz überhaupt gefragt wird. Dies wird in drei Abschnitten geschehen, die – ausgehend von der Theorie der Turing-Maschinen – die Differenz von natürlicher und künstlicher Intelligenz präziser zu erfassen suchen. In einem zweiten Teil werden – wiederum in einem Dreischritt – die so gewonnenen Einsichten für anthropologische, ethische und theologische Fragestellungen fruchtbar gemacht.

I. Künstliche Intelligenz

1. Turing-Maschinen

Die Frage nach der künstlichen Intelligenz lässt sich formulieren als die Frage, ob mit Hilfe formaler, nach mathematischen Regeln operierender Systeme Funktionen menschlicher Intelligenz nachgebildet und auf künstlich hergestellten Maschinen realisiert werden können, so dass diesen Maschinen selbst Intelligenz zugeschrieben werden muss. Doch insofern damit behauptet wird, dass auf diese Weise Eigenschaften kontrolliert nachkonstruiert werden können, die traditionellerweise dem Menschen zugeschrieben werden, meldet sich darin der Anspruch, dass nun auf objektive und an der unumstößlichen Gewissheit mathematischer Erkenntnis teilhabende Weise nachvollzogen wird, was denn Intelligenz überhaupt ausmacht. Theorie und Pragmatik der technischen Systeme, so der weitergehende Schluss, umfassen auch Theorie und Praxis menschlicher Intelligenz, ja bringen ihr Verständnis gerade wegen ihrer mathematisierten Form aus dem bloßen Herumtappen in den sicheren Gang einer Wissenschaft.

So wie die klassische Mechanik die Bewegungsgesetze zurückführte auf ideale störungsfreie Trägheitsbewegungen, so führt die Theorie der künstlichen Intelligenz das Denken zurück auf die ideale Maschine. Als das

Modell einer idealen Denk-Maschine gilt die universelle Turing-Maschine¹, eine diskret steuernde Maschine mit unendlicher Speicherkapazität, die elektromechanisch realisiert oder auch mit heutigen Computern simuliert werden kann – natürlich ohne die ihrer Theorie zugrunde gelegten idealen Eigenschaften wie unendliche Speicherkapazität und unbegrenzt zur Verfügung stehende Zeit. Es lässt sich zeigen, dass alle algorithmischen, rekursiven Verfahren auf der universellen Turing-Maschine simuliert werden können.

Inzwischen sind viele Computer und Roboter realisiert, die der Theorie der Turing-Maschinen gehorchen und algorithmisch gesteuert Daten verarbeiten, so dass sie zu differenziertem Problemlösungsverhalten in der Lage sind und »intelligent« auf bestimmte Aufgabenstellungen reagieren können. Dazu gehören komplexe Entscheidungsprozesse, wie sie in Schachcomputern oder auch in Geräten zur chemischen Analyse und medizinischen Diagnose zu finden sind und wie sie sich von der frühen Theorie des *General Problem Solver* (GPS)² bis zu heutigen regelgeleiteten Expertensystemen entwickelt haben. Solche computergesteuerten Entscheidungssysteme erhalten ihre Informationen durch Rezeptoren, die optische, akustische, olfaktorische und andere Reize in verarbeitbare Daten übersetzen und so zu überaus differenziertem Problemlösungsverhalten in der Lage sind. Aber auch immer ausgefeiltere Kommunikationssysteme, linguistische Analyseprogramme und selbständig agierende und auf ihre Umwelt reagierende Roboter wären hier zu nennen.

Auch aus theologischer Sicht dürfte es einigermaßen unproblematisch sein, solchen Maschinen Intelligenz zuzuschreiben und diese als »künst-

¹ Alan M. Turing (1912–1954), britischer Mathematiker, Logiker und Computerwissenschaftler, entwickelte 1936–1937 die Theorie der nach ihm benannten Turing-Maschinen (Alan M. TURING, On Computable Numbers, With An Application to the Entscheidungsproblem [1937], in: DERS., Intelligence Service. Schriften, Bernhard DOTZLER/Friedrich KITTLER, Berlin 1987, 18–60). Eine Turing-Maschine verfügt über eine endliche Anzahl diskreter innerer Zustände und einen äußeren Speicher in Form eines Magnetbandes. Dieses prinzipiell unendlich lange Band ist in Felder unterteilt, die an einem Lese-/Schreibkopf vorbeilaufen. Dieser ist in der Lage, ein Zeichen, das in das jeweilige Feld eingetragen ist, zu lesen, ein Zeichen in ein leeres Feld zu schreiben oder ein vorhandenes Zeichen zu löschen. Der Lese-/Schreibkopf ebenso wie die schrittweise Vorwärts- bzw. Rückwärtsbewegung des Bandes werden nach Maßgabe der inneren Zustände der Maschine gesteuert, die außerdem festlegen, wie von einem inneren Zustand in einen anderen übergegangen werden soll, je nachdem, welche Zeichen vom Band gelesen werden. Turing war in der Lage zu zeigen, dass Turing-Maschinen im Prinzip alles tun können, was jede beliebige Rechenmaschine können muss, welcher Bauart sie auch sei. Turing zeigte auch, dass es kein endliches Entscheidungsverfahren gibt, das für alle möglichen Turing-Maschinen und alle möglichen Eingaben entscheidet, ob die Turing-Maschine mit dieser Eingabe nach endlich vielen Schritten anhält oder endlos weiterläuft, das so genannte Halteproblem.

² Vgl. dazu Herbert A. SIMON/Allen NEWELL, Informationsverarbeitung in Computer und Mensch, in: Walther Ch. ZIMMERLI/Stefan WOLF (Hg.), Künstliche Intelligenz. Philosophische Probleme, Stuttgart 1994, 112–145, besonders 135 ff.

liche Intelligenz« zu apostrophieren³. Denn diese Maschinen erbringen Leistungen, die wir als »intelligent« bezeichnen würden, wenn ein Mensch sie erbrächte. Insofern sie aber kein Mensch, sondern eine Maschine erbringt, ist es sinnvoll, dieser Maschine eine nicht natürlich entstandene, sondern eben eine *künstliche*, also durch von Menschenhand konstruierte Maschinen erbrachte Intelligenz zuzuschreiben. *Intelligenz* ist wohl auch nicht missverstanden, wenn man sie als einen regelgeleiteten Prozess versteht, der in Verfolgung vorgegebener Ziele auf externe, in symbolisierter Form prozessierte Daten mit angemessenen Reaktionen antwortet und womöglich die auf die erteilte Antwort erfolgenden Reaktionen noch einmal in Bezug auf die verfolgten Ziele analysiert, um die ihn steuernde Regel zu optimieren. In diesem Sinne behauptet der Ausdruck »künstliche Intelligenz« nicht zu viel.

Die prinzipielle und weiterführende Frage, die sich nun stellt, ist allerdings die, ob die Prinzipien, die diesen Maschinen zugrunde liegen, nur die *notwendigen* oder aber auch die *hinreichenden* Bedingungen von Intelligenz *überhaupt* darstellen. Dass Letzteres in der Tat der Fall sein könnte, legt sich im Rahmen einer naturwissenschaftlich orientierten und als Naturalismus bezeichneten Grundanschauung nahe. Unter *Naturalismus* sei in diesem Zusammenhang die These verstanden, dass alle realen Systeme, seien es Sterne, Planeten, Steine, Pflanzen, Tiere oder Menschen, »aus mehr oder weniger komplexen Anordnungen von materiellen Konstituentien« bestehen, »die durch eine kleine Zahl von dynamischen Wechselwirkungen zusammengehalten werden«⁴, und zwar genau denjenigen Wechselwirkungen, die die mathematisierten Naturwissenschaften zu beschreiben in der Lage sind. Auch das Verhalten großer und komplexer Systeme in seiner ganzen Vielfalt ließe sich dann zumindest prinzipiell zurückführen auf die Eigenschaften der elementaren Bausteine und ihre Wechselwirkungen.

Nun ist für alle kognitiven Prozesse menschlicher Intelligenz dasjenige System bekannt, das von Rezeptoren symbolisierte Daten prozessiert und regelgeleitet in angemessenes Verhalten umsetzt, nämlich unser Gehirn. Einzelne mentale Operationen lassen sich heute bestimmten Gehirnarealen zuordnen, emotive und kognitive Leistungen werden durch Schädigungen von Gehirnformationen spezifisch beeinträchtigt, und auch die einheitlichen Bausteine des Gehirns, die Nervenzellen, sind bekannt und in einigen Aspekten ihrer Funktionalität analysiert. Die Neuronen scheinen dabei im Prinzip wie elektro-chemische Schalter zu funktionieren und sollten durch berechenbare Funktionen darstellbar sein. Darüber hinaus ist der ganze kognitive Apparat im Rahmen eines naturalistischen Weltbilds als Produkt

³ So auch Eilert HERMS, Künstliche Intelligenz, in: DERS., Gesellschaft gestalten. Beiträge zur evangelischen Sozialethik, Tübingen 1991, 284–295, hier: 288.

⁴ Bernulf KANITSCHIEDER, Vorwort, in: DERS./Franz J. WETZ (Hg.), Hermeneutik und Naturalismus, Tübingen 1998, V.

evolutionärer Entwicklung und Optimierung anzusehen, so dass die ihm extern vorgegebenen Ziele in umwelt-adäquatem Verhalten und überlebensfördernder Effizienz bestehen, sowohl im Sinne der individuellen Fitness des Individuums wie auch der Gesamtfitness der Gattung.

Damit legt sich der Schluss nahe, dass auch der menschliche kognitive Apparat nichts anderes ist als ein General Problem Solver und damit ein Spezialfall einer Turing-Maschine. Genau auf dieser Linie argumentierte Turing selber, um die Intelligenz von Maschinen als Paradigma von Intelligenz an sich zu behaupten. Zwar gesteht Turing zu, dass das Nervensystem mit Sicherheit keine diskrete, also schrittweise arbeitende Maschine sei⁵. Dennoch hält er fest, dass Gehirne dieser Klasse von diskreten Systemen offensichtlich sehr nahe stehen, »und es scheint allen Grund für die Annahme zu geben, sie könnten auch so beschaffen sein, daß sie gänzlich darunter fallen, ohne irgendeine Veränderung ihrer wesentlichen Eigenschaften«⁶. Deshalb sollten alle Zustände, in denen sich das Gehirn bei Entscheidungsoperationen befindet, vollkommen durch entsprechende Zustände einer Turing-Maschine repräsentiert werden können: »Jedem Geisteszustand des Rechnenden entspricht ein ›*m*-Zustand‹ der Maschine.«⁷

Daran anschließend hat Alan Turing den bekannten und später nach ihm so genannten *Turing-Test* entwickelt, der die Frage entscheiden soll, wann einem maschinellen System Intelligenz zugeschrieben werden kann. Es handelt sich um ein Spiel, an dem ein Mensch, eine Maschine und ein Fragesteller teilnehmen. Alle drei finden sich in getrennten Räumen, und der Fragesteller weiß nicht, in welchem Raum sich die Maschine und in welchem sich der Mensch befindet. Seine Aufgabe ist es, Mensch und Maschine zu befragen, um herauszufinden, in welchem Raum die Maschine steht. Die Maschine soll so programmiert werden, dass sie den Fragesteller zu täuschen versucht. Wenn er nach einer ausreichenden Anzahl von Durchgängen in nicht wesentlich mehr als der Hälfte der Versuche den Menschen richtig identifizierte, hat die Maschine den Test bestanden, so dass ihr ein dem menschlichen analoges Denkvermögen zuzuschreiben ist.

2. Zur Differenz zwischen natürlicher und künstlicher Intelligenz

Mit seinem Testverfahren für die Zuschreibung von Intelligenz hat Turing im Grunde ein unschlagbares Argument geliefert: Sage mir klar und deutlich, worin der Unterschied zwischen Mensch und Maschine in ihrem

⁵ Alan M. TURING, *Computing Machinery and Intelligence* (1950), in: DERS., *Intelligence Service. Schriften* (s. o. Anm. 1), 148–182, hier: 171.

⁶ Alan M. TURING, *Intelligent Machinery* (1969), in: aaO., 82–113, hier: 87. Vgl. auch aaO., 96: »Die Schaltkreise, die in elektronischen Rechenmaschinen verwendet werden, scheinen die wesentlichen Eigenschaften von Nerven zu haben.«

⁷ TURING, *On Computable Numbers* (s. o. Anm. 1), 43.

Verhalten besteht, und ich baue eine Maschine, die auch diese Differenz noch simuliert. Denn sobald du etwas klar und deutlich beschreiben kannst, kann es algorithmisch, also regelgeleitet nachvollzogen werden.

Ausgehend von dieser schlecht von der Hand zu weisenden Argumentation, dass, sobald wir eine Differenz im Verhalten von Mensch und Maschine nicht bloß intuitiv behaupten oder raten, sondern bestimmt angeben können, worin sie phänomenologisch besteht, diese Differenz auch wieder simulierbar ist, ergeben sich zwei mögliche Argumentationsmuster. Zum *einen* kann man sie als ein Argument für den Reduktionismus in Anschlag bringen und nun behaupten, die natürliche Intelligenz sei im Grunde auch nichts anderes als die künstliche. Turing hat seinen Test in dieser Absicht verstanden. Reduktionisten können sich durch die prinzipiell zum Scheitern verurteilte Forderung, einen phänomenalen Unterschied klar und deutlich zu benennen, der dann selbst nicht wieder programmierbar wäre, berechtigt fühlen, die Behauptung, dass Computer ebenso denken können wie Menschen, so lange aufrecht zu erhalten, wie das *eigentliche* Wesen von natürlicher Intelligenz nicht aufgedeckt ist. Doch sobald phänomenale Kriterien zu dessen Identifizierung klar und deutlich bestimmt werden, kann man sogleich eine neue Maschine postulieren, die auch diese vorgeblichen Wesenseigenschaften natürlicher Intelligenz noch simuliert.

Zum *anderen* kann man Turings Testverfahren aber auch als Argument gegen den Reduktionismus verstehen. Gegner des Reduktionismus können immer darauf hinweisen, dass die Differenz Mensch und Maschine eben im Nicht-Empirischen liegen muss, in dem, was nicht klar und deutlich bestimmt werden kann, sondern intuitiv erfasst werden muss. Eine diskret arbeitende Turing-Maschine, so kann argumentiert werden, ist dann prinzipiell nicht in der Lage, durch Simulation das einzuholen, was nur intuitiv und nicht nach formalen Regeln erkannt werden kann und was nur natürlicher Intelligenz eigen ist, die deshalb mehr und anderes sein muss, als was funktional durch Maschinen modelliert werden kann.

Beide Argumentationsstrategien führen dazu, dass der jeweilige Protagonist immer gewinnt. Daraus ist ersichtlich, dass der Turing-Test für unsere Fragestellung eigentlich nichts austrägt, sondern in eine aporetische Patt-Situation führt, die von woanders her und mit Argumenten entschieden werden muss, die das Verhältnis von empirisch feststellbarem Verhalten und aus unserer Selbsterfahrung und Selbstdeutung stammenden, vor-empirischen Evidenzerfahrungen betreffen. Der Turing-Test ist lehrreich in der Hinsicht, dass er darauf aufmerksam macht, dass, wenn es überhaupt eine Differenz zwischen Systemen künstlicher Intelligenz und dem Menschen gibt, diese darin bestehen muss, dass natürliche Systeme *nicht vollständig* über ihre regelgeleitete und empirisch feststellbare Funktionalität bestimmt sind.

Das kann man auch mit folgendem Gedankengang illustrieren. Eine Turing-Maschine ist bloße Sklavin ihres Programms, und sie ist durch die

Funktionalität dieses Programms hinreichend beschrieben. Turing hatte in seiner Arbeit eine konkrete Maschine entworfen und zu Grunde gelegt, doch nur, um an ihr die eigentliche Theorie rekursiver Funktionen zu entwickeln. Die technische Ausführung ist dazu unerheblich. Statt eines Magnetbandes, auf das mit Hilfe eines Magnetkopfes beschrieben, gelesen und gelöscht wird, ließe sich auch ein Papierband verwenden, auf das ein Stift schreibt, ein entsprechender Sensor liest und ein Radierer löscht, ohne dass sich die Formel änderte, die die Funktionalität dieser Maschine beschreibt.

Die Eigenart einer Turing-Maschine, so würden wir heute sagen, besteht in ihrer Software, die Hardware ist beliebig. Wir können deshalb ihr Programm, ihre Funktion von gänzlich verschiedenen Maschinen realisieren lassen. Darunter sind auch solche Maschinen, denen wir intuitiv gerade keine Fähigkeit des Denkens zuschreiben würden. So könnte jede Turing-Maschine als mechanisch-pneumatische Maschine realisiert werden. Doch dass Ventilen und Kolben künstliche Intelligenz im Sinne von Denken zukommt, wird man nicht so leicht behaupten wollen⁸.

Umgekehrt kann man die Funktionen einer Turing-Maschine aber auch durch ein bewusstes denkfähiges Wesen wie einen Menschen ausführen lassen, ohne dass dieser versteht, was er tut. Dies ist das bekannte *Argument des chinesischen Zimmers* von John Searle⁹. In diesem Gedankenexperiment sitzt ein Mensch, der des Deutschen, aber nicht des Chinesischen mächtig ist, in einem abgeschlossenen Zimmer und hat vor sich eine Aufstellung von Regeln, wie welche deutschen Worte durch welche chinesischen Schriftzeichen, die ihm vorgefertigt zur Verfügung stehen, zu ersetzen

⁸ Die These der sog. *starken künstlichen Intelligenz*, dass alle Formen menschlichen Bewusstseins und menschlicher Intelligenz statt mit Neuronen auch mit Hilfe von elektronischen Schaltkreisen erzeugt werden können, gründet auf einer solchen Unterscheidung von Hardware und Software, doch ist fraglich, ob diese Differenz immer konsequent durchdacht wird. Beim Vergleich von Mensch und Computer soll es unerheblich sein, ob neuronale Strukturen oder elektronische Schaltkreise das Programm ausführen. Entscheidend sei die Funktionalität, denn nur diese sei empirisch feststellbar. Sobald aber unterschiedliche Systeme ununterscheidbare Effekte hervorbringen, seien beiden die gleichen Eigenschaften in Bezug auf diese Effekte zuzuweisen. Wenn wir jedoch sagen, das Entscheidende sei einzig und allein das Programm, dann wären Denken, Bewusstsein etc. der Theorie der Turing-Maschinen zufolge im Prinzip auch auf *mechanischen* Maschinen simulierbar, denn jeder elektronische Computer kann in seiner Funktionalität mit mechanischen Maschinen wiederum simuliert werden. Die Suggestionkraft der starken künstlichen Intelligenz beruht sicher mit auf der geheimnisvollen Kraft der Elektrizität, die mit Geisthaftigkeit assoziiert ist und die mit ungeheurer Geschwindigkeit nahezu verlustfrei arbeitet, damit aber für uns unbeobachtbar bleibt. Intuitiv erscheint es uns aber nicht beliebig, ob Neuronen, elektronische Schaltkreise oder pneumatische Ventile »Denkprozesse« durchführen. Wir trauen dem ersten Baustein mehr zu als dem zweiten und diesem wiederum mehr als dem dritten. Die »Hardware« erscheint in der Theorie der Turing-Maschinen beliebig, in der Realität jedoch urteilen unsere im Umgang mit uns selbst und mit unserer Umwelt geschulten Intuitionen anders.

⁹ John SEARLE, Geist, Gehirn, Computer, in: ZIMMERLI/WOLF (s. o. Anm. 2), 232–265.

sind. Wann immer man einen Zettel mit einem deutschen Text von außen durch einen Briefkastenschlitz ins Zimmer schiebt, wird man nach einiger Zeit von dem Menschen im Zimmer eine Zusammenstellung von chinesischen Schriftzeichen zurück erhalten, die eine chinesische Übersetzung des deutschen Textes darstellen. Vorausgesetzt, die rein mechanisch auszuführenden Regeln sind gut und der Mensch verfolgt sie gewissenhaft, wird auch die Übersetzung gut sein und man könnte von außen den Eindruck bekommen, im Zimmer sitze jemand, der des Chinesischen mächtig ist. Doch der die Regeln Ausführende tut dies rein mechanisch, ohne das geringste Verständnis für das, was er da tut, auch wenn er von seinen geistigen Fähigkeiten her prinzipiell dazu in der Lage wäre, Chinesisch zu beherrschen. Daraus ergibt sich der Schluss: Ein Programm, eine Funktion, ein regelgesteuertes Verhalten muss nicht selbst denken und verstehen können, um zu funktionieren.

Wenn also Maschinen, die aus Bauelementen bestehen, denen wir die Fähigkeit wirklichen Denkens und bewusster Intelligenz grundsätzlich absprechen würden, in ihrer Funktionalität jedem Computer gleichzusetzen sind, und wenn umgekehrt ein denkendes Wesen alles das ausführen kann, was ein Computer tut, ohne ein bewusstes Verständnis für die Sache zu entwickeln, dann ist deutlich, was künstliche Intelligenz von natürlicher Intelligenz unterscheidet: Künstliche Intelligenz befolgt Regeln, natürliche Intelligenz versteht Bedeutungen. Auch wenn das wahrnehmbare Verhalten zwischen beiden Systemen nicht grundsätzlich und in bestimmten Situationen wie dem Turing-Test vielleicht gar nicht phänomenal zu unterscheiden wäre, ist diese Differenz festzuhalten. Jeder rein formal operierenden Maschine ist dann eben *nur* künstliche, das heißt nicht bedeutungstragende und verstehende Intelligenz zuzuschreiben. Oder umgekehrt, natürliche intelligente Systeme sind nicht bloß formal operierende Maschinen.

3. *Semantische Offenheit*

Dass aber Turing-Maschinen schon aus prinzipiellen Gründen nicht alles können, zeigt sich in einer von Turing selbst entdeckten Eigenart der Turing-Maschinen, dem so genannten Halteproblem¹⁰. Es besagt, dass man zwar für jedes Problem eine Maschine bauen kann, die es zu lösen versteht, dass es aber keine Maschine gibt, die *jedes* Problem zu lösen in der Lage ist. Dass ein dem Halteproblem für Turing-Maschinen äquivalentes Problem für alle formalen Systeme gilt, die mindestens so stark sind, dass sie die Arithmetik ausdrücken können, zeigten schon die beiden Unvollständigkeitssätze von Kurt Gödel (1906–1978), die dieser 1931 veröffentlichte¹¹.

¹⁰ S. o. Anm. 1.

¹¹ Kurt GÖDEL, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, in: Monatshefte für Mathematik und Physik 38 (1931), 173–198.

Der erste besagt, nicht-formal formuliert, dass es in jedem entsprechend starken System (informal) wahre Aussagen gibt, die mit den Mitteln des Systems nicht formal bewiesen werden können. Daraus folgt mit einigen Zusatzüberlegungen der zweite Satz, dass, wenn ein entsprechendes System als widerspruchsfrei angenommen wird, diese Widerspruchsfreiheit unabweisbar ist. Kein formales System kann deshalb sowohl widerspruchsfrei als auch vollständig sein, es enthält, wenn es widerspruchsfrei ist, notwendigerweise unentscheidbare Aussagen.

Doch machen die Gödel'schen Theoreme auch auf ein Phänomen aufmerksam, das uns zu dem von uns geltend gemachten Unterschied zwischen mechanisch-maschinellem und natürlich-menschlicher Intelligenz zurückführt. Gödels Beweise machen deutlich, dass der Begriff der mathematischen *Wahrheit* nicht innerhalb desselben Systems mitkonstruiert werden kann. Zwar hat Gödel gerade gezeigt, dass die Differenz zwischen Objekt- und Metasprache durch sein Verfahren der Gödelisierung aufgehoben werden kann. Der Begriff der Beweisbarkeit kann mit Mitteln des Systems selbst dargestellt werden. Doch gerade dann, wenn das der Fall ist, lässt sich nicht mehr sicherstellen, dass das System vollständig und widerspruchsfrei ist. Es ist demnach auch nicht die bloße Selbstbezüglichkeit, die durch selbstreferentielle Schließungen die Unvollständigkeit des Systems hervorruft, es ist die Tatsache, dass ein System seine eigenen Aussagen semantisch bewertet, wenn es sich selbst die Wahrheitsfrage stellt. Turing-Maschinen sind *semantisch geschlossene* Systeme, denen ein Wahrheitsbegriff durch ihre Programme und die Diskretheit der Programmschritte vorgegeben wird. Natürlich verstehende Systeme sind *semantisch offene* Systeme, die die Differenz von Objekt- und Metasprache in sich noch einmal erzeugen und so immer aufs Neue nach Wahrheit fragen und sich für Wahrheitsbedingungen *entscheiden* müssen.

In diesem Sinne hat Gödel selbst die Konsequenzen seiner Entdeckung interpretiert. Der Fortschritt der Mathematik wie überhaupt aller menschlichen Erkenntnis ist für ihn als immer neue und weiter fragende Sinnklärung zu verstehen, die prinzipiell nicht auf eine endgültige, abgeschlossene Theorie zulaufen kann. Das menschliche Denken kann und muss sich immer wieder selbst in Richtung auf die Gewinnung neuer möglicher Wahrheit und Erkenntnis hin überschreiten: »Es zeigt sich nämlich, daß bei einem systematischen Aufstellen der Axiome der Mathematik immer wieder neue und neue[re] Axiome evident werden, die nicht formallogisch aus den bisher aufgestellten folgen [...] eben dieses Evidentwerden immer neuerer Axiome auf Grund des Sinnes der Grundbegriffe ist etwas, was eine Maschine nicht nachahmen kann.«¹²

¹² Vgl. den Vortragsentwurf *The modern development of the foundations of mathematics in the light of philosophy* von 1961 in: Kurt GÖDEL, *Collected Works*, vol. III, hg. v. Solomon FEFERMAN, New York/Oxford 1995, 372–387, hier: 384.

Ein rein syntaktischer Begriff von Wahrheit in einer semantisch geschlossenen Sprache führt zu Widersprüchen und Unvollständigkeit. Wahrheit, so ist dann umgekehrt zu sagen, ist ein Beziehungsbegriff, der nur dann Sinn macht, wenn er relativ zu und in Unterscheidung von einem syntaktischen, einem algorithmisch rekursiv arbeitenden System formuliert wird. Und diese Relationalität von Wahrheit ist nicht finit, sie kann sozusagen nicht durch eine umfassende, abschließende Theorie wiederum formalisiert und mit Hilfe einer Rechenvorschrift noch einmal von der Theorie eingeholt werden. Zwar ist es an sich möglich, jede Semantik mit Mitteln der interpretierten Systeme abzubilden und so zu einem Teil des formalen Systems zu machen. Doch gilt dann wiederum, dass sich für ein jedes solches System ›wahre‹ Sätze aufstellen lassen, die mit Mitteln des Systems nicht mehr beweisbar sind. Der Schluss legt sich jedenfalls nahe und ist vielfach gezogen worden, »daß die Quellen des menschlichen Intellekts nicht vollständig formalisiert wurden und daß dies auch in Zukunft nicht möglich ist«¹³.

II. Künstliche Intelligenz in theologischer Perspektive

Aus dem bis jetzt Vorgeführten lassen sich in mancherlei Hinsicht Konsequenzen und Perspektiven entwickeln, die für Theologie und Ethik von einiger Bedeutung sind. Dazu sollen im Folgenden einige relevante Überlegungen angestellt werden, die sich drei Fragestellungen zuordnen lassen: 1. Was bedeutet künstliche Intelligenz für unser Menschenbild? 2. Welche ethischen Fragen und Problembereiche in Bezug auf Systeme künstlicher Intelligenz schließen sich an? 3. Was bedeuten diese Überlegungen für ein theologisches Gottesbild und die Erkenntnisleistung bzw. Erkenntnisgrenzen von Theologie?

1. Mensch und Maschine

Zunächst einmal ist eine Erklärung des Bewusstseins im Rahmen eines evolutionistischen Weltbildes im Sinne eines Fitness steigernden Problemlösungsapparates als unvollständig anzusehen. Seit der Etablierung der Evolutionstheorie als dem durchgängigen Paradigma zur genetischen Erklärung biologischer Systeme wurde auch der kognitive Apparat des Menschen in Analogie zu rein physiologischen Körperorganen als Instrument zur Sicherung eines Überlebensvorteils gesehen. Als eine Stimme unter vielen mag hier der Physiker Ludwig Boltzmann zu Wort kommen: »Das Ge-

¹³ Ernest NAGEL/James R. NEWMANN, *Der Gödelsche Beweis*, München⁶2001, 99.

hirn betrachten wir als den Apparat, das Organ zur Herstellung der Weltbilder, welches sich wegen der großen Nützlichkeit dieser Weltbilder für die Erhaltung der Art entsprechend der Darwinschen Theorie beim Menschen geradeso zur besonderen Vollkommenheit herausbildete, wie bei der Giraffe der Hals, beim Storch der Schnabel zu ungewöhnlicher Länge.«¹⁴

Das ist zunächst als genetische Erklärung nicht falsch, doch als hinreichende Beschreibung unzulänglich. Der Überlebensvorteil der eigentlichen Verstehensdimension des Bewusstseins wird dann jedenfalls obsolet, wenn man es auf bloße Steuerungs- und Regelungsfunktionen reduziert. Dazu hätte ein bewusstloser zentraler kybernetischer Apparat ausgereicht. Doch ein solcher Apparat wäre nicht in der Lage gewesen, wirkliches Verstehen und semantische Offenheit zu erzeugen. Das Fragen nach und Verstehen von Bedeutungen aber ist eine notwendige Bedingung für die Entstehung von Sprache, Kommunikation, Sozialität und Wissenschaft, aber auch von Religion.

Grundlegend für die verstehende Wahrnehmung von Welt ist dabei die Differenz zwischen wahrnehmendem Subjekt und wahrgenommener Wirklichkeit. Nur durch die Unhintergebarkeit dieser Unterscheidung, die ihrerseits nicht wieder durch das erkennende System rekonstruierend eingeholt werden kann, ist zum Beispiel auch der Aufwand unseres Wahrnehmungsapparates erklärbar, eine einheitliche Erfahrungswelt zu konstruieren, in der selbst der blinde Fleck unseres Gesichtsfelds noch wegetuschiert wird und uns eine kohärente Welt gegenübertritt. Als bewusste Wesen haben wir eine *Welt*, in der wir sind und der wir zugleich gegenüberstehen, ein General Problem Solver zur Regelung von Prozessen bräuchte nur eine Menge von *Informationen*. Und ebenso *haben* wir nicht bloß ein Selbstmodell als interne formale Repräsentation unseres Weltbezugs¹⁵, sondern *sind* wir ein wirkliches Ich. Wir brauchen das Gegenüber von Welt und Subjekt, um die semantische Geschlossenheit eines nur formalen Problemlösens durchbrechen zu können.

Unsere Subjektivität ist allerdings wiederum enthalten in einer anderen, sie mit umfassenden Hierarchieebene, in der auch die Differenz zwischen Subjekt und Welt wieder eingebettet ist, in die soziale Kommunikationsgemeinschaft von Subjekten. Ein Mensch wird erst am Du zum Ich, lautet die schlichte Form, in die Martin Buber diese Einsicht gebracht hat. Wir müssen erst durch die Beziehungen, mit denen wir aufwachsen, und

¹⁴ Ludwig BOLTZMANN, Über die Frage nach der objektiven Existenz der Vorgänge in der unbelebten Natur, in: DERS., Populäre Schriften, Leipzig 1905, 162–187, hier: 179.

¹⁵ So versteht Thomas Metzinger z. B. menschliche Gehirne als »General Problem Solvers« und vergleicht sie mit Flugsimulatoren, die jedoch zugleich unser Selbstbewusstsein als ein internes Selbstmodell zusammen mit den Szenarien der Außenwelt generieren: »Menschliche Gehirne simulieren den Piloten gleich mit« (Thomas METZINGER, Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation, Paderborn 1993, 243).

durch die Sprache, mit der wir uns in Auseinandersetzung mit anderen entwerfen, lernen, dass wir ein Subjekt sind. Wir brauchen den anderen, um ein Selbstverhältnis, um Subjektivität und Personalität entwickeln zu können.

Doch auch die Pluralität der Semantiken, der Interpretationen, die wir kommunikativ generieren, kann nicht wiederum in einem Super-System vollständig zusammengefasst werden, ohne sie zu zerstören. Denn auch und gerade unser kommunikatives Handeln kann nicht rein funktional formalisiert werden, denn das Ich des anderen bleibt mir auf direkte Weise unzugänglich. Das meint nicht, dass es nicht funktionalen Zwecken dient, es lässt sich nur nicht auf solche reduzieren, ohne seine Eigenart der Generierung und Zuschreibung von Bedeutung und Sinn zu verlieren. Auch soziales und kommunikatives Handeln lässt sich nicht in bloß instrumentelles Handeln auflösen.

Insofern wir also nicht nur funktional und regelgeleitet operieren, gehört eine auch theologisch höchst relevante Kategorie mit zur Charakterisierung von Personalität und Menschsein: die in technischen Zusammenhängen zumeist nur als Störung zu verstehende Kategorie der *Unterbrechung*¹⁶. Wir können über uns erschrecken, wenn wir bemerken, dass wir in unserem Leben nur ein Programm abspulen, dass wir in eine dumpfes, regelgeleitetes *Funktionieren* abgeglitten sind. Ob auch eine auf Optimierung hin programmierte kybernetische Steuerungsmaschine über sich erschrecken und aufgrund eines empfundenen Defizits in ethischer oder geistiger Hinsicht sich selbst unterbrechen und fragen könnte: Was machst du da eigentlich? Warum funktionierst du bloß? Es ist jedenfalls ein wichtiger Aspekt von Religion, dass sie solche Unterbrechungen herbeiführen will, im Gottesdienst, in Gebet und Meditation und anderem mehr. Aber auch weltlich macht sich ein spontaner Einfall oder eine intuitive Wahrnehmung als Unterbrechung von Routinen bemerkbar.

Bloße Funktionalität jedenfalls wäre Grundlage eines dürftigen Menschenbildes. Sie ignoriert die letzte Unverfügbarkeit, in der wir uns als handelnde Subjekte gegenüberstehen, und sie ignoriert, dass wir uns Beziehungen verdanken, die nicht formal beherrschbar sind, ohne dass man sie zerstört. An diese Grundstruktur des Menschseins knüpft die theologische Rechtfertigungslehre an: Wir sind nicht unser eigenes Werk, auch nicht das Machwerk eines uns hergestell habenden Schöpfers. Und wir dürfen und sollen weder unserem herstellenden Handeln unterwerfen, was durch Beziehung konstituiert wird, noch umgekehrt das, was durch Beziehung konstituiert ist, durch hergestellte Objekte ersetzen.

¹⁶ Vgl. Eberhard JÜNGEL, Wertlose Wahrheit. Christliche Wahrheitserfahrung im Streit gegen die Tyrannie der Werte, in: DERS., Wertlose Wahrheit. Zur Identität und Relevanz des christlichen Glaubens (Theologische Erörterungen III), Tübingen (1990) ²2003, 90–109, hier: 100 ff.

Die Gefahr der Maschinenmetapher ist dann die, dass angesichts der Tatsache, dass der Mensch in vielen Zusammenhängen immer mehr und immer effektiver verzweckt wird, die Behauptung der Simulation humaner Eigenschaften den Trend verstärkt, dass extern vorgegebene Effizienzkriterien den Maßstab bilden für gelingendes Dasein. Das Problem in der Auseinandersetzung mit Robotik und künstlicher Intelligenz scheint mir deshalb nicht so sehr zu sein, dass Roboter und Computer sich *uns* immer mehr angleichen, so dass wir in unserer Einzigartigkeit gekränkt würden, sondern dass wir uns genötigt sehen oder auch nur versucht sind, uns *ihnen* immer mehr anzugleichen. Maschinen, so sahen wir, arbeiten heteronom nach extern vorgegebenen Effizienzkriterien, die sie passiv und, wenn meine Interpretation richtig ist, auch ohne Verstehen erfüllen. Diese Verzweckung scheint auf den ersten Blick zu kongruieren mit einer Sicht der Evolution, die als Überleben des Tüchtigeren verstanden wird und sich dadurch als das Gesetz der Natur schlechthin imponiert. Und zugleich ist dies das Gesetz der globalisierten Ökonomie, dem immer mehr Bereiche menschlicher Existenz unterworfen werden. »Wo Natur und Gesellschaft gemeinsam mitwirken, ein auf bloßes Effizienzdenken reduziertes Menschenbild zu zementieren, setzt sich eine weltanschauliche Grundstimmung durch, die die These plausibel erscheinen lässt, auch der Mensch sei schließlich nichts anderes als ein Roboter.«¹⁷ Die nach Effizienzgesichtspunkten vorgehende Durchrationalisierung der Gesellschaft ist schon so weit fortgeschritten, dass es oft keinen Unterschied mehr macht, wenn man den Menschen durch eine Maschine ersetzt. Hier scheinen mir die eigentlichen Bedenken gegen die Einebnung des Unterschiedes zwischen natürlicher und künstlicher Intelligenz zu liegen, und nicht in der allzu abstrakten Furcht davor, als »Krone der Schöpfung« entthront zu werden. Damit kommt nun die eigentlich ethische Dimension unseres Themas in den Blick.

2. Ethische Fragestellungen

Für die folgenden Erörterungen müssen wir noch einmal zu der Frage zurückkommen, ob und wie wir denn künstliche Intelligenz herzustellen in der Lage sind. Wir hatten als eine Konsequenz aus den Gödelschen Theoremen festgehalten, dass bei rein formal arbeitenden Entscheidungssystemen ein semantisches Verstehen im Rahmen der eigenen Prozesse nicht anzunehmen ist. Doch wenn das nicht durch algorithmische Programmierung direkt hergestellt werden kann, dann kann es vielleicht, und so hatte

¹⁷ Hans-Dieter MUTSCHLER, Ist der Mensch ein Roboter?, in: Matthias KOSSLER/Reinhard ZECHER (Hg.), Von der Perspektive der Philosophie. Beiträge zur Bestimmung eines philosophischen Standpunkts in einer von den Naturwissenschaften geprägten Zeit, Hamburg 2002, 291–308, hier: 306.

auch schon Turing argumentiert, durch *lernende* Systeme erzeugt werden¹⁸, die in einen sozialen und kommunikativen Kontext mit Menschen eingebettet sein könnten und denen möglicherweise funktionale Einheiten zugrunde liegen, die nicht vollkommen diskret arbeiten und deshalb keine exakte Turing-Maschine darstellen.

Dass dies möglich sein könnte, lässt sich sicher nicht ausschließen. Mit anderer Hardware und offeneren Programmstrukturen mag es möglich sein, Systeme zu entwickeln, denen wir nicht Bewusstsein einprogrammiert haben, sondern bei denen so etwas wie Bewusstsein im Sinne von semantischem Verstehen im Laufe von Lernprozessen sich einstellen könnte. Dann aber stellt sich nicht mehr nur die Frage, ob und wie wir solche Maschinen konstruieren *können*, sondern auch, ob wir das *wollen*.

Auch von solchen Maschinen würde gelten, was wir über menschliches Bewusstsein selbst gesagt haben, dass ihr innerer verstehender Zustand von außen in direkter Weise unzugänglich wäre. Wieder hat schon Turing dies festgehalten: »Ein wichtiges Kennzeichen einer lernenden Maschine ist, daß ihr Lehrer oft reichlich wenig von dem weiß, was genau in ihr vorgeht, wenn er auch bis zu einem gewissen Grad dennoch in der Lage sein mag, das Verhalten seines Schülers vorauszusagen.«¹⁹ Wenn wir das Vermögen von semantischer Differenz und Offenheit auf einer Maschine realisieren wollen, dann müssen wir wohl auch zulassen, dass die Maschine selbständig die sie leitenden Zwecke modifiziert. Sie könnte sich möglicherweise eigene Zwecke setzen und gesetzte Zwecke aufheben. Ein Roboter, dem wir Bewusstsein und verstehende Intelligenz zuschreiben könnten, müsste sich also in irgendeiner Weise sein Gesetz selbst geben können. Er wäre damit wohl auch in der Lage, sich zu unterbrechen und sich zu fragen, was er da eigentlich macht und warum. Er könnte streiken. Er könnte ›entarten‹. Kurz, ein solches Wesen würde dem Begriff der Maschine widersprechen, die ja auf externe Zwecke hin optimierte Funktionen ausführen soll.

Die Frage ist, ob wir *solche* Roboter bauen wollen. Doch warum bauen wir Systeme mit möglichst großer künstlicher Intelligenz eigentlich überhaupt? Drei Gründe scheinen mir vor allen Dingen bedeutsam:

1. Wir können objektivierende, berechnende Instrumente zur Entscheidungsvorbereitung gut gebrauchen, gerade weil wir selbst nicht rein instrumentell funktionieren (Expertensysteme).

2. ›Intelligent‹ gesteuerte mechanische Maschinen können vielfältige Aufgaben übernehmen und Dinge tun, die uns körperlich schwer fallen, bei

¹⁸ Vgl. TURING, Computing Machinery (s. o. Anm. 5), 177: »Warum sollte man nicht versuchen, statt ein Programm zur Nachahmung des Verstandes eines Erwachsenen eines zur Nachahmung des Verstandes eines Kindes herzustellen? Unterzöge man dieses dann einem geeigneten Erziehungsprozeß, erhielte man den Verstand eines Erwachsenen.«

¹⁹ AaO., 181.

denen aber komplexe Entscheidungen notwendig sind (Roboter in schwieriger Mission im Weltraum, bei der Verbrechensbekämpfung, in Produktionsprozessen), und die uns intellektuell schwer fallen, weil sie in ihrer Komplexität für uns heuristisch schwer zu bewältigen sind (Suchmaschinen im Internet).

3. Wir wollen wissen, »wie es geht«, und damit auch darauf schließen können, wie wir selbst funktionieren. Bei den Robotern, mit denen wir direkt und ausschließlich menschenähnliches Verhalten nachkonstruieren, dürfte die Neugier auf die konstruierte Vernunft mit eine Rolle spielen. Und da hinein mischen sich oft auch Visionen, die uns künstliche Systeme geradezu als die besseren Menschen vorführen.

Einige Sätze von Hans Moravec, einem Roboterbauer, mögen dies illustrieren: »Ich sehe diese Maschinen als unsere Nachkommen. Im Augenblick glaubt man das kaum, weil sie eben nur so intelligent sind wie Insekten. Aber mit der Zeit werden wir das große Potential erkennen, das in ihnen steckt. Und wir werden unsere neuen Roboterkinder gern haben, denn sie werden angenehmer sein als Menschen. Man muß ja nicht all die negativen menschlichen Eigenschaften, die es seit der Steinzeit gibt, in diese Maschinen einbauen. Damals waren diese Eigenschaften für den Menschen wichtig. Aggressionen etwa brauchte er, um zu überleben. Heute, in unseren großen zivilisierten Gesellschaften machen diese Instinkte keinen Sinn mehr. Diese Dinge kann man einfach weglassen – genauso wie den Wesenszug der Menschen, daß sie ihr Leben auf Kosten anderer sichern wollen. Ein Roboter hat das alles nicht. Er ist ein reines Geschöpf unserer Kultur und sein Erfolg hängt davon ab, wie diese Kultur sich weiterentwickelt. Er wird sich also sehr viel besser eingliedern als viele Menschen das tun. Wir werden sie also mögen und wir werden uns mit ihnen identifizieren. Wir werden sie als Kinder annehmen – als Kinder, die nicht durch unsere Gene geprägt sind, sondern die wir mit unseren Händen und mit unserem Geist gebaut haben.«²⁰

Auch wenn solche Visionen reine Fiktion sind und hier vieles nüchtern abgewartet werden kann, ist es durchaus angezeigt, Leitlinien für die Beantwortung der Frage zu entwickeln, welche intelligenten Maschinen wir denn zu welchen Zwecken haben wollen. Die folgenden drei Punkte scheinen mir dabei unverzichtbar zu sein.

1) Ein Hauptproblem ist die Frage nach der *Delegation von Verantwortung* an Maschinen. Effizienzdenken und die Frage nach der Verantwortlichkeit im Schadensfalle fördern sicher die Versuchung, persönliche Verantwortung zu delegieren an Expertensysteme. Eine deutliche Grenze aber setzen hier *qualifizierte ethische Entscheidungen* mit einer Tragweite, die menschliches Leben direkt betreffen. Es verbietet sich, ein Expertensystem zum Beispiel über das Abschalten einer Herz-Lungen-Maschine entscheiden zu lassen oder ein Waffensystem über den Einsatz von lebensvernichtenden Waffen oder ein Strategieprogramm über politische Entscheidungen, denn hier spielen nicht-formalisierbare Entscheidungen über die Bedeutung von Leben und Menschsein die entscheidende Rolle.

²⁰ Hans MORAVEC, *Computer übernehmen die Macht: Vom Siegeszug der künstlichen Intelligenz*, Hamburg 1999, 136.

2) Aus einem ähnlichen Grund ist auch die *Jurisprudenz* als eine Kunst zu würdigen, die humane Urteilskraft und Verantwortlichkeit erfordert, nicht das Resultat eines mechanisch anwendbaren Verfahrens sein kann und deshalb nicht an Expertensysteme delegiert werden darf. Es ist sicher kein Zufall, dass der Kant'sche Begriff der ›Urteilskraft‹ ursprünglich aus der Rechtssphäre stammt. Wer Urteile fällt, braucht diese ›Urteilskraft‹, die nicht vollständig formalisierbar ist und auch allein in persönlicher Verantwortung stehende Billigkeits-Erwägungen mit einschließt.

Aber auch viele andere Gebiete menschlichen Erkennens und Handelns kommen ohne solche Urteilskraft nicht aus. Sie vollziehen sich zwar nicht chaotisch, sondern durchaus nach Regeln, diese aber sind so offen, dass sie nur in Bezug auf persönliche Verantwortung und Interpretation Gestalt gewinnen, wie zum Beispiel in der *Pädagogik*, *Ästhetik* oder literarischen *Hermeneutik*. Auch hier ist Skepsis geboten, wenn formalisierte Maschinenintelligenz zu dominieren beginnt.

3) Ein weiterer Punkt ist die Frage nach der *sozialen Komponente* der Computer-Mensch-Beziehung. Festzuhalten ist, dass eine Interaktion mit Systemen künstlicher Intelligenz nicht die Interaktionen zwischen Menschen ersetzen kann und darf. Die Unterhaltung mit einem vielleicht bald realisierten Pflegeroboter bei der Versorgung kranker oder alter Menschen darf die menschliche Begegnung nicht ersetzen.

Gegen diese Möglichkeiten gilt es, nicht einfach zu behaupten, dass nicht sein kann, was nicht sein darf, sondern die Frage wach halten, ob das sein muss, was sein kann. Es gilt also auch die Eigendynamik der technischen Entwicklungen ihrerseits zu unterbrechen und die Frage »Wozu?« zu stellen, damit nicht nur unter Absehung von Endzwecken Teilmechanismen optimiert werden, die dann zusammen mit politischen und ökonomischen Zwängen problematische Trends freisetzen. Dann ist es sehr bedenklich, erscheint unter technischen und ökonomischen Gesichtspunkten aber auch als überflüssig, bewusste künstliche Wesen erzeugen zu wollen, die nicht als bloße Maschinen, sondern als ›lebensartig‹ anzusehen wären. Und unsere Neugier, wie wir denn funktionieren (aber wir funktionieren ja Gott sei Dank auch nicht nur!), würde erst recht nicht befriedigt, denn wir könnten diese Maschinen wohl auch nicht besser verstehen als uns selbst.

3. Glauben und Wissen

Werfen wir zunächst noch einmal einen Blick zurück auf unsere Ausführungen zum Menschenbild und entfalten wir den entsprechenden Begriff von Gott als Schöpfer. Wenn der Mensch nicht ein evolutionär optimierter Problemlösungsapparat ist und in seinem Sein und Wesen nicht bloß externe Zwecke zur Geltung kommen, sondern der Mensch als semantisch offenes Wesen sich selbst und seine Welt zu verstehen sucht und

nur als relationales Wesen existieren kann, dann ist das Bekenntnis zu Gott als seinem Schöpfer nicht der Hinweis auf einen transmundanen Ingenieur oder Programmierer. Der sich in den ungeheuren Räumen und Zeiten des Kosmos vollziehende Entwicklungsprozess der Schöpfung ist nicht Ausdruck eines herstellenden Handelns. Wenn in einer besonders ausgezeichneten Nische aus überaus verhaltenen Anfängen und aus einer überschwänglichen Fülle von Gestalten auch der Mensch als ein Wesen entsteht, das nach sich selbst fragt, dann ist er ebenso wenig wie alle anderen Kreaturen das Machwerk eines göttlichen Konstrukteurs. Wir sind, wie alle Schöpfung, um unserer selbst willen da. Und im Blick auf unser Dasein und Sosein ist Gott nicht als unser Hersteller zu preisen, sondern als die Quelle, der Grund und der Antrieb der Fülle von Möglichkeiten, der seiner Schöpfung gegenwärtig bleibt und uns als solchermaßen entstehende und vergehende Geschöpfe bejaht und in die Eigentlichkeit unserer beziehungsreichen Existenz ruft.

Das heißt dann aber auch, dass der Gedanke eines unabhängig von der Wirklichkeit der sich vollziehenden Schöpfung über das Ganze des Geschehens »voll informierten Schöpfers«²¹ abzuweisen ist. Oder genauer, und also mit Gödel gesagt, aus der Annahme, dass der Inbegriff aller Wahrheiten in einem einzigen System erfasst werden könnte, lässt sich ein Widerspruch herleiten, wenn vorausgesetzt ist, dass das System widerspruchsfrei ist. An Wittgensteins bekanntem Diktum, dass die Welt die Gesamtheit der Tatsachen ist, ist deshalb dies problematisch, dass die wahren Tatsachen keine Gesamtheit, keine widerspruchsfreie und vollständige Totalität darstellen²². Es ist in einer semantisch offenen Sprache kein Subjekt widerspruchsfrei denkbar, das »einen epistemisch ausgezeichneten externen Standpunkt einnehmen« könnte, »von dem aus es zu allen wahren Aussagen bzw. Sachverhalten epistemischen Zugang besitzt«²³. Der theologische Begriff von Gottes »Wahrheit« dürfte deshalb wesentlich vom Begriff der »Wahrhaftigkeit« Gottes her zu bestimmen sein und nicht von der Vorstellung des Besitzes einer Totalität von Satz Wahrheiten²⁴.

Deshalb ist umgekehrt auch kein Gottesbeweis möglich, denn ein Begriff unüberbietbarer Totalität ist für uns unerreichbar. Jeder Beweis ist nur relativ zu einem System von Voraussetzungen zu führen. Einen vorausset-

²¹ Manfred EIGEN/Ruth WINKLER, *Das Spiel*, München ⁵1983, 224.

²² Vgl. Patrick GRIM, *The Incomplete Universe*, Cambridge, Mass./London 1991, 2f. und 6.

²³ Elke BRENDEL, *Wahrheit und Wissen*, Paderborn 1999, 165.

²⁴ Das dürfte auch in Joh 14,6 gemeint sein, wo der Gottessohn von sich sagt, er sei der Weg, die Wahrheit und das Leben. Das Semitische kennt kein Abstraktum, und so gehören alle drei Begriffe zusammen und sind auf der gleichen semantischen Ebene anzuordnen. Der Bezugspunkt unseres Denkens, Glaubens und Meinens sind dann nicht die ewigen, im Verstande Gottes residierenden Vernunftwahrheiten, sondern es ist die Wahrhaftigkeit Gottes, die wir erfahren in seiner Schöpfung, in seiner Rechtfertigung, auf die wir uns verlassen in konkreten Situationen des Lebens, auf die wir hoffen in Bezug auf die Bejahung unseres Daseins.

zungslosen Beweis gibt es nicht. Das Absolute ist von relativen Voraussetzungen her gerade unerreichbar. Die Behauptung, dass Gott existiert, ist deshalb zwar eine notwendige Implikation und Voraussetzung des Glaubens, zugleich aber ist die Frage nach ihrer Wahrheit *remoto fide* eine in absolutem Sinne unentscheidbare Frage.

Der Glaube verlangt aber auch nicht nach Beweisen, nicht nach Vollständigkeit und Widerspruchsfreiheit von Argumenten oder Systemen. Denn auch das beste Wissen könnte den Glauben unter Bezug auf den derzeitigen Stand des Wissens nur wahrscheinlich machen – doch »das Christentum als das Wahrscheinliche – [...] dann ist das Christentum abgeschafft«²⁵. Eher verlangt der Glaube seinerseits nach einer *Unterbrechung*, nach einer Suspension der diskursiven, der formalisierten Vernunft, die Raum gibt dafür, vertrauensvoll zu leben. Es geht der Religion gerade um die Dimension menschlichen Daseins, die sich nicht beweisend einfordern oder formal herstellen lässt und die doch fundamental ist für unsere beziehungsreiche Existenz: »Sympathie, Vertrauen, wechselseitige Achtung und Anerkennung, Freundlichkeit, Gütigkeit, oder, im Kontext christlicher Glaubenskommunikation, Liebe oder Barmherzigkeit.«²⁶ Religion versucht diese Dimension des Lebens zugänglich und kommunizierbar zu machen, ohne sie selbst wieder zu formalisieren und so unter Kontrolle zu bringen. Das versuchen allenfalls Aberglauben und Magie. Aber auch die weltliche und wissenschaftliche Vernunft, das scheinen mir die Gödel'schen Theoreme nahe zu legen, nimmt immer schon Vertrauen in Anspruch. Verständnis ist ohne vorgängiges Einverständnis nicht möglich, eine Rationalität ohne Vertrauen ist unvernünftig, eine Wahrheit ohne Wahrhaftigkeit ist leer.

Damit ergibt sich, was auch der Wissenschaftstheoretiker und Philosoph Wolfgang Stegmüller als Konsequenz aus der Analyse der Logik und der Gödel'schen Theoreme festgestellt hat: »Eine ›Selbstgarantie‹ des menschlichen Denkens ist, auf welchem Gebiete auch immer, ausgeschlossen. Man kann nicht vollkommen ›voraussetzungslos‹ ein positives Resultat gewinnen. Man muß bereits an etwas glauben, um etwas anderes rechtfertigen zu können. Mehr könnte sinnvollerweise nur dann verlangt werden, wenn wir die Endlichkeit unseres Seins zu überspringen vermöchten. Aber der archimedische Punkt außerhalb unserer endlichen Realität bleibt, zumindest für uns, eine Fiktion.«²⁷

Die christliche Theologie ist deshalb primär zu verstehen als eine Heuristik zur Orientierung im Offenen und Endlichen, nicht als abschlussfähige

²⁵ Søren KIERKEGAARD, Papirer, Kopenhagen ²1968 ff., V, Bd. X-4, A 633.

²⁶ Johannes FISCHER, Pluralismus, Wahrheit und die Krise der Dogmatik, in: ZThK 91 (1994), 487–539, hier: 496.

²⁷ Wolfgang STEGMÜLLER, Metaphysik Skepsis Wissenschaft, Berlin/Heidelberg/New York ²1969, 307.

Theorie über transzendente Gegenstände. Und auch für die Theologie gilt, »auf Letztbegründungsversuche und damit auf die Suche nach unhintergehbaren Gegebenheiten, letzten Identitäten und unhinterfragbaren außertheologischen Sachverhalten ganz zu verzichten«. Auch ihr stehen keine zwingenden letzten Gründe, »sondern nur relative, interne und lokale Kriterien«²⁸ zur Verfügung. Damit aber steht sie als Erkenntnisbemühung nicht isoliert da im Zusammenhang verstehenden Denkens des Menschen überhaupt, mit dem er auch und besonders nach sich selbst fragt. Und so mag ein gewagter Satz am Ende unserer Ausführungen stehen, der aus unverdächtigem Munde oder besser unverdächtigter Feder, nämlich noch einmal von Wolfgang Stegmüller stammt: »Denken, Lieben und Beten haben also *etwas* gemeinsam.«²⁹

SUMMARY

The article asks for the conditions and restraints of artificial intelligence and reflects on its relevance for humans in a theological perspective. It first presents an account of Alan M. Turing's analysis of computational intelligence and uses it to specify the difference between (up to date) artificial and natural intelligence. It is claimed with reference to Kurt Gödel's Incompleteness Theorems that natural intelligence is distinguished from currently possible forms of artificial intelligence by semantic openness. The second part of the article develops consequences for human self-understanding in difference to machines, answers the question which robots to which purposes we should and should not build, and asks for the relation between faith and reason with respect to the incompleteness of formal knowledge and description.

²⁸ Ingolf U. DALFERTH, Subjektivität und Glaube. Zur Problematik der theologischen Verwendung einer philosophischen Kategorie, in: NZSth 36 (1991), 18–58, hier: 49.

²⁹ STEGMÜLLER (s. o. Anm. 27), 456.