



Philosophie der Digitalität

6. Sitzung, 31.5.2019

Thema heute:
Künstliche Intelligenz

Thema heute: Künstliche Intelligenz



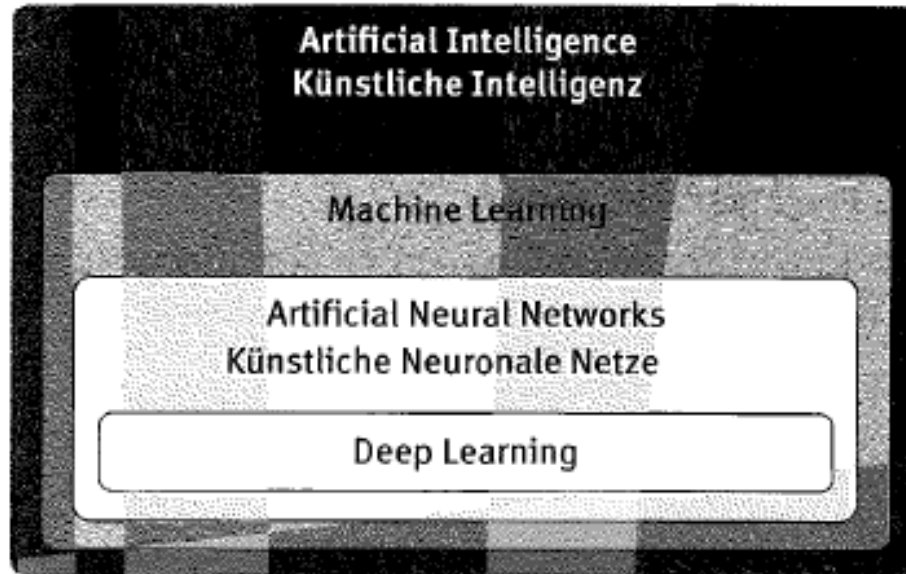
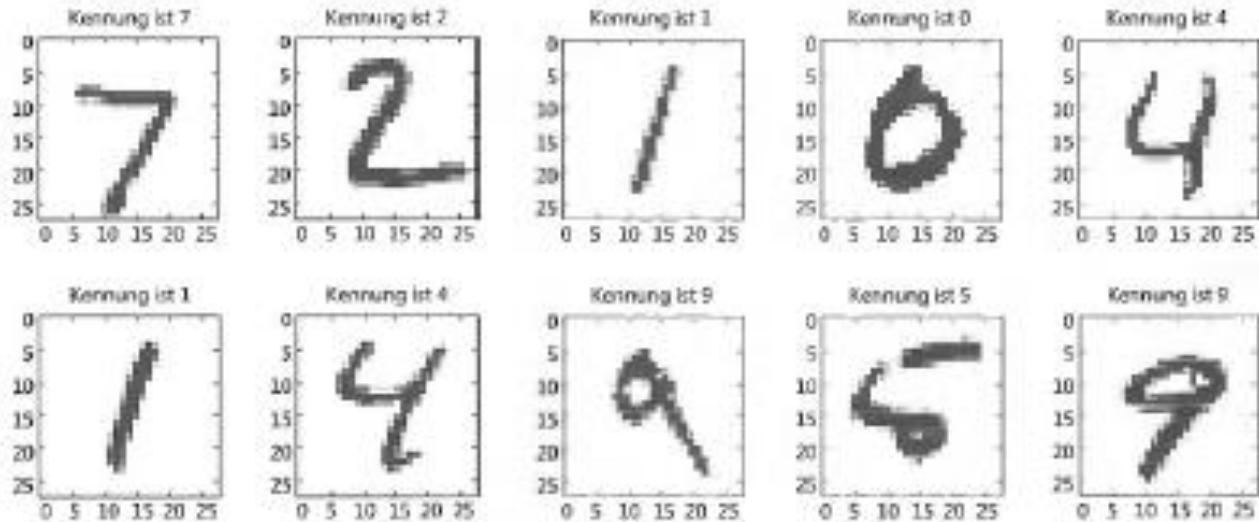


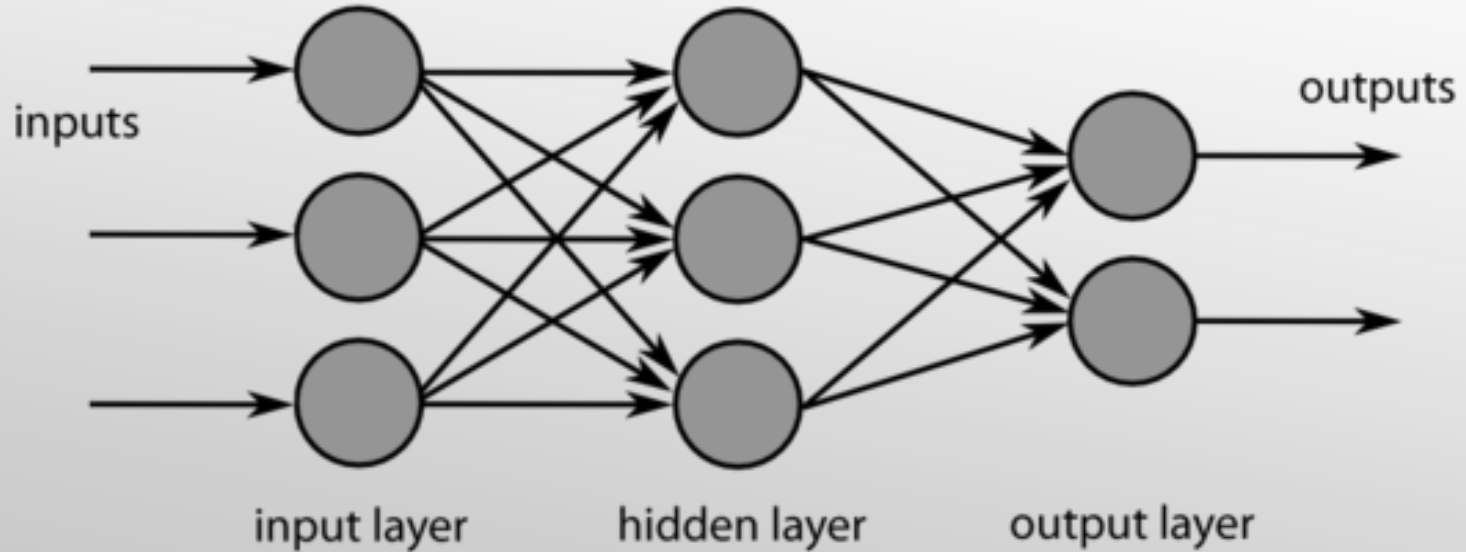
Abbildung 1.11 Begriffszwiebel für künstliche Intelligenz

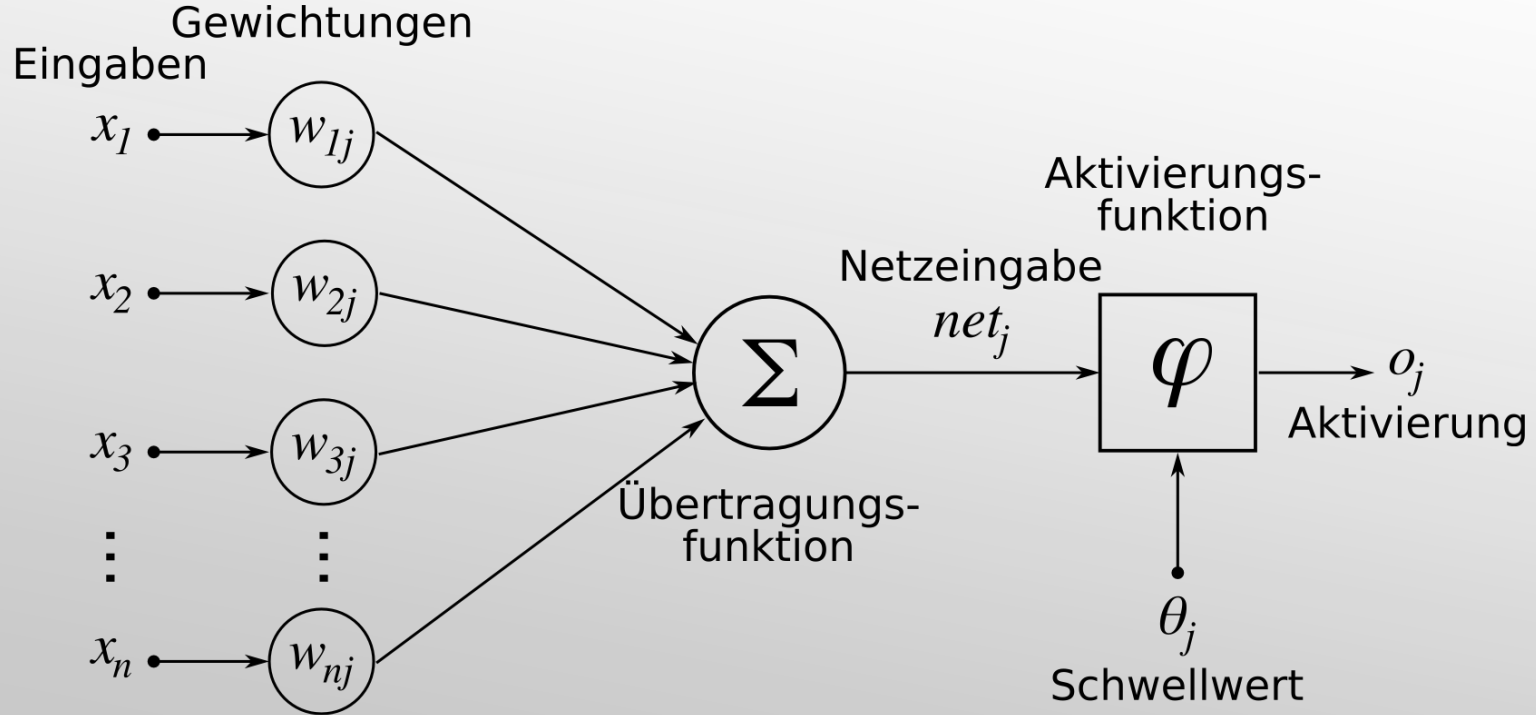


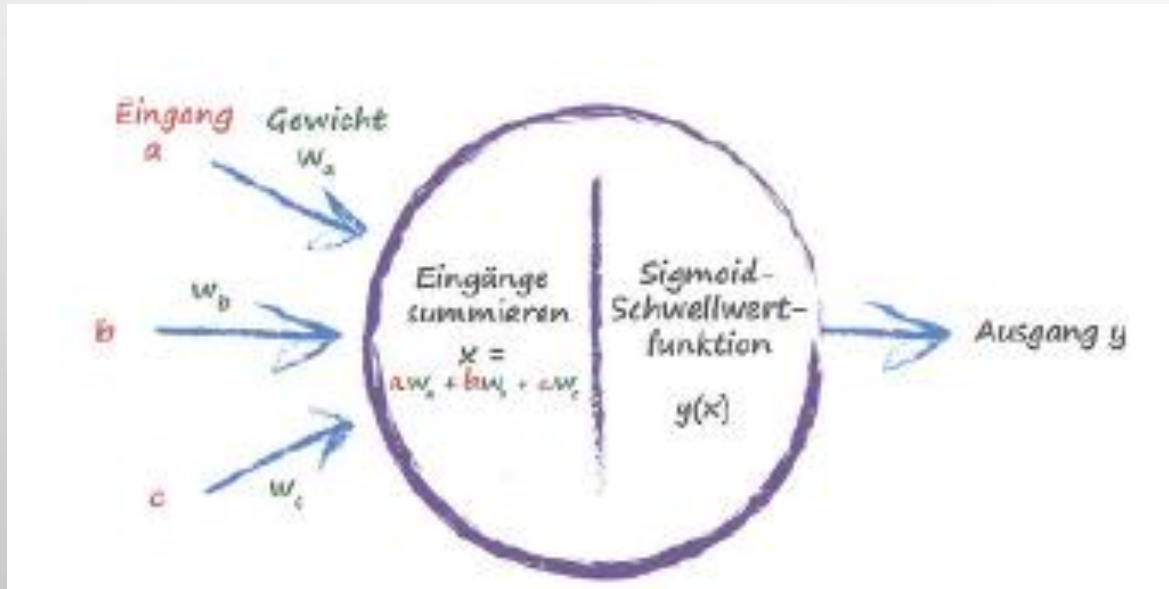
„Drehen wir nun den Spieß um und tauschen wir die Rolle mit dem Computer! Sehen Sie sich die folgenden Bilder an und versuchen Sie, zu erkennen, was sie enthalten:



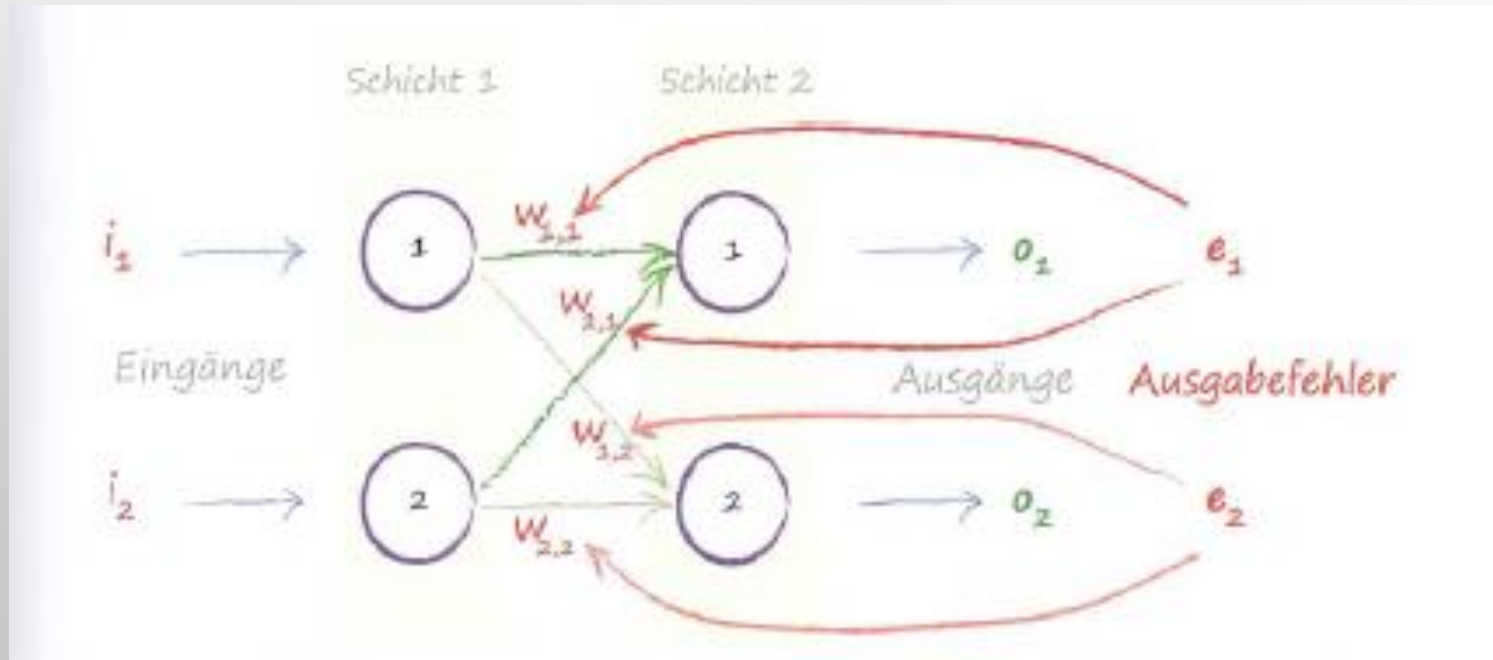
Abbildung 1-1: Bilderkennung – einfacher für den Computer oder für den Menschen?



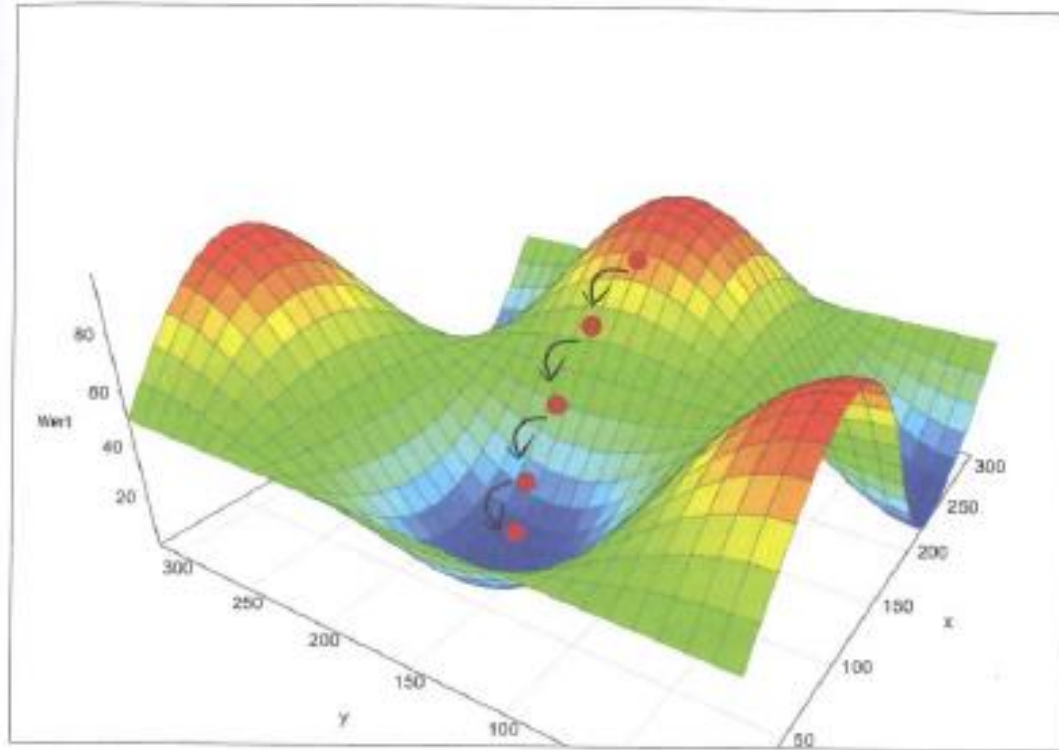




Fehlerrückmeldung („backpropagation“)



Gradienten-Verfahren





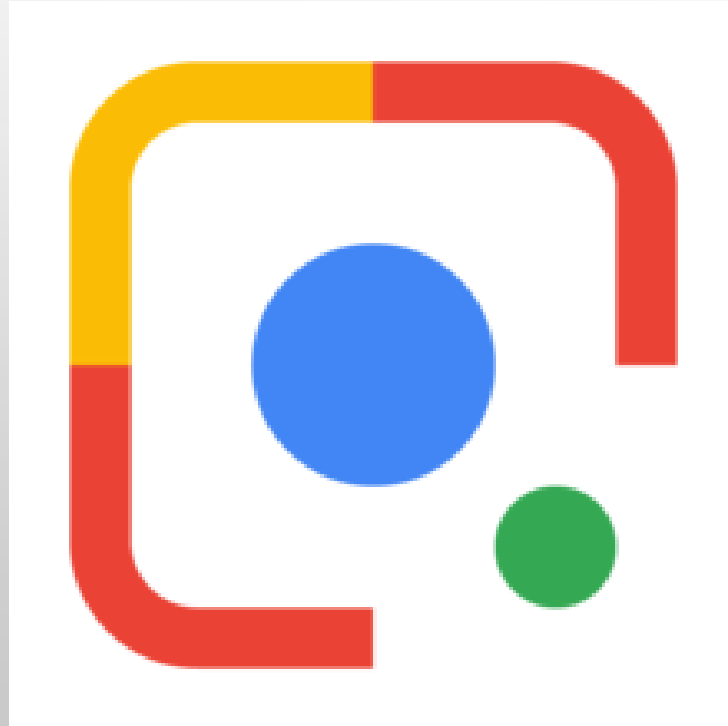
Epistemische Phänomene bzw. Akte

- Klassifizieren bzw. kategorisieren
- Erkennen
- Lernen
- Merken
- Vernetzen
- Erinnern
- Wissen
- Verstehen
- Denken
- Begreifen
- Vorstellen
- Entscheiden -> Handeln



„Urteilstkraft überhaupt ist das Vermögen, das Besondere als enthalten unter dem Allgemeinen zu denken. Ist das Allgemeine (die Regel, das Prinzip, das Gesetz) gegeben, so ist die Urteilstkraft, welche das Besondere darunter subsumiert [...] bestimmend. Ist aber nur das Besondere gegeben, wozu sie das Allgemeine finden soll, so ist die Urteilstkraft bloß reflektierend.“

(Kritik der Urteilstkraft, AA V, 179)







Intelligenz bedeutet Problemlösungskompetenz.

Problem: Das Erkenntnisinteresse wird der künstlichen Intelligenz von außen („heteronom“) vorgegeben.

Kann für KI etwas aus eigenem Interesse problematisch sein?

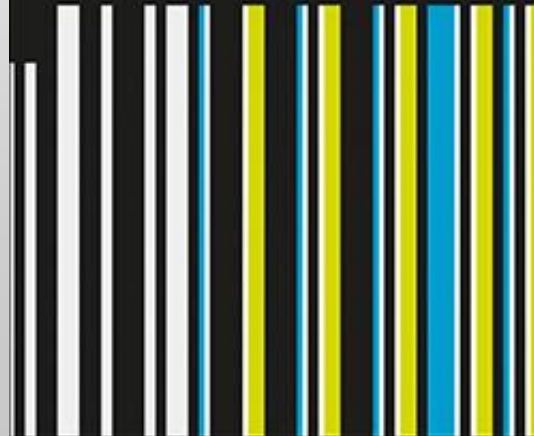
Es scheint, dass nur lebendige Systeme eigenes („autonomes“) Problembewusstsein haben.



[Was bedeutet das alles?]

Thomas Ramge
Mensch und Maschine
Wie Künstliche Intelligenz
und Roboter unser Leben
verändern

Reclam





Leitfragen:

- Was bedeutet "künstliche Intelligenz" (KI)?
- Worin liegen die Grenzen, wo die Potenziale der KI?
- Inwiefern können wir Maschinen als Subjekte oder Akteure auffassen?
- Inwiefern können wir ihre Operationen als "Erkennen", "Wissen", "Entscheiden", "Denken" oder gar "Handeln" bezeichnen?
- Worin liegt die moralische Problematik von KI



„Künstliche Intelligenz (KI) erlebt gerade ihren Kitty-Hawk-Moment. Den KI-Forschern ergeht es wie den Pionieren des Motorflugs. Jahrzehntlang sind sie immer wieder gescheitert, nach vollmundigen Ankündigungen ein ums andere Mal abgestürzt. Doch dann gelingt den Brüdern Wright in Kitty Hawk im US-Staat North Carolina der Durchbruch. Die Technologie hebt ab und plötzlich klappt, was vor wenigen Jahren nur eine großsprecherische Behauptung war. Seit rund drei Jahren können Computerprogramme menschliche Gesichter deutlich zuverlässiger erkennen als die meisten Menschen. Bei der Diagnose bestimmter Krebszellen sind Rechner heute schon genauer als die besten Ärzte der Welt – geschweige denn als durchschnittliche Ärzte in einem Provinzkrankenhaus. Computer schlagen den Menschen nicht nur im intuitiven Brettspiel Go, seit Januar 2017 ist amtlich: Sie bluffen auch besser als die besten Pokerspieler der Welt. Bei der japanischen Versicherung Fukoku Mutual prüft das IBM-System Watson die Rückerstattungsansprüche der Versicherten. Bei Bridgewater, dem größten Hedgefonds der Welt, entscheiden Algorithmen nicht nur über Investitionen. Ein mit umfangreichen Mitarbeiterdaten gefüttertes System wird zum Robo-Boss: Es kennt die wahrscheinlich beste Geschäftsstrategie, die beste Zusammensetzung eines Teams für bestimmte Aufgaben, und es gibt Empfehlungen zu Beförderungen und Entlassungen.“



„KI ist die nächste Stufe der Automatisierung. Schweres Gerät erledigt schon seit Langem die schmutzige Arbeit. Fertigungsroboter wurden seit den 1960er Jahren immer geschickter. IT-Systeme halfen bis dato vor allem bei den Routineschleifen der Wissensarbeit. Sie erleichtern die Buchhaltung, rechnen im Auftrag des Menschen oder verarbeiten Texte. Doch mit Künstlicher Intelligenz treffen jetzt Maschinen komplexe Entscheidungen, die bisher nur Menschen treffen konnten. Oder genauer formuliert: Wenn Datengrundlage und Entscheidungsrahmen stimmen, entscheiden KI-Systeme besser, schneller und billiger als LKW-Fahrer, Sachbearbeiter, Verkäufer, Ärzte, Investmentbanker oder Personal-Manager.“



„Wer die Chancen und Risiken einer neuen Technologie erkunden möchte, muss zunächst die Grundlagen verstehen. Er muss verständliche Antworten auf die Fragen finden: Was ist Künstliche Intelligenz überhaupt? Was kann sie heute und in absehbarer Zeit? Und welche Fähigkeiten muss der Mensch weiterentwickeln, wenn Maschinen immer intelligenter werden und den Menschen zu überflügeln sich anschicken? Nach Antworten auf diese Fragen sucht dieses Buch.“



Der Tesla fährt im Autopilot-Modus mit 130 Kilometern pro Stunde auf der linken Spur der Autobahn. Auf der rechten Spur fahren mehrere Lastwagen mit 90 km/h. Der Tesla nähert sich der Kolonne. Der letzte LKW setzt links den Blinker und will überholen. Der Autopilot muss eine komplexe Entscheidung treffen. Soll der Tesla mit gleicher Geschwindigkeit weiterfahren oder gar beschleunigen, um auf jeden Fall den Laster passiert zu haben, bevor dieser eventuell die Fahrbahn wechselt? Sollte er hupen, um den LKW-Fahrer zu warnen? Wäre das in dieser Situation erlaubt? Oder soll der Tesla bremsen und dem LKW das Überholmanöver höflich erlauben, auf Kosten der Reisegeschwindigkeit, aber zugunsten der Sicherheit? Wobei bremsen freilich nur dann sicher wäre, wenn von hinten kein von Testosteron gesteuerter Sportwagenfahrer mit zwei Metern Abstand drängelt.



Ein altes Bonmot in der KI-Forschung lautet: Maschinen fällt leicht, was Menschen schwerfällt, und umgekehrt. Autofahren mit seinen abertausend kleinen, aber dennoch komplexen Entscheidungssituationen während einer Fahrt war Computern bisher nicht möglich. Warum ändert sich das gerade? Abstrakt gesprochen lautet die Antwort: Weil aus Daten lernende Software in Verbindung mit steuerungsfähiger Hardware den Dreischritt von Erkennen, Erkenntnis und Umsetzung in eine Handlung immer besser beherrschen.



Aus Situation, Regeln und Erfahrung leitet das System eine Erkenntnis ab, nämlich die beste Möglichkeit aus vielen errechenbaren Szenarien, einen Unfall zu vermeiden und dennoch zügig voranzukommen. Im Kern handelt es sich dabei um eine kognitive Entscheidung, also die Auswahl einer Handlungs-Option unter vielen. Die beste Lösung des Problems ist das Ergebnis einer Wahrscheinlichkeitsrechnung, in die viele Variablen einfließen.



Zur Natur künstlich intelligenter Systeme gehört, dass sie die Auswirkungen ihrer Entscheidungen messen und die Ergebnisse in künftige Entscheidungsfindung einbeziehen. Sie entscheiden mit Feedback-Schleifen. Wenn der Tesla in der beschriebenen Situation einen Unfall baut, funkt es dieses Feedback zurück in den Zentralrechner, und alle anderen Teslas werden nach dem nächsten Software-Update in vergleichbaren Situationen (hoffentlich) deutlich defensiver fahren. Verzeichnet eine KI-Software zur Kreditvergabe zu viele Ausfälle, die es selbst autorisiert hat, wird es die Kriterien für kommende Kreditbewerber verschärfen. Wenn eine Erntemaschine die Rückmeldung erhält, zu viele unreife Äpfel zu pflücken, wird es beim nächsten Durchgang besser entscheiden können, ob ein bestimmtes Verhältnis der Färbung aus rot und grün auf der Apfeloberfläche ausreicht, um zuzugreifen. In dieser Fähigkeit, die eigenen Berechnungen und die Einordnung ihrer Ergebnisse selbstständig zu verbessern, liegt der wesentliche Unterschied zwischen Künstlicher Intelligenz und klassischen IT-Systemen. Die Selbstkorrektur ist im System eingebaut.



Seit den ersten Großrechnern in den 1940er Jahren bedeutete das Programmieren eines Computers: Der Mensch bringt einer Maschine mühsam ein theoretisches Modell bei. Dieses Modell besteht aus bestimmten Regeln, welche die Maschine anwenden kann. Wenn die Maschine die passenden Daten für bestimmte Aufgaben oder Fragen zugefüttert bekommt, kann sie diese dann in der Regel schneller, genauer, zuverlässiger und günstiger lösen als der Mensch. Auch das war im Ergebnis immer beeindruckend, aber im Kern wurde bei klassischer Programmierung existierendes Wissen aus den Köpfen der Programmierer in eine Maschine übertragen. Dieser technische Ansatz hat eine natürliche Grenze. Ein großer Teil unseres Wissens ist implizit.



Der Begriff Künstliche Intelligenz ist umstritten, seit die Computer-Pioniere um Marvin Minsky ihn 1956 auf ihrer berühmten Dartmouth-Konferenz prägten. Dabei ist die Wissenschaft sich bis heute nicht einmal darin einig, was menschliche Intelligenz genau ausmacht. Kann so ein Begriff dann überhaupt für Maschinen taugen? Diskussionen um Künstliche Intelligenz driften schnell zu sehr grundsätzlichen Fragen ab wie: Ist Denken ohne Bewusstsein möglich? Sind Maschinen bald intelligenter als Menschen, und bilden sie die Fähigkeit aus, sich selbst immer intelligenter zu machen? Und können sie dann unter Umständen ein Selbstbild, ein Bewusstsein und eigene Interessen entwickeln? Die Bewegung der Transhumanisten – eine wilde Mischung aus Techno-Utopisten oft mit esoterischem Einschlag – glaubt oder hofft, dass Menschen und Maschinen irgendwann verschmelzen, sich menschliches Bewusstsein auf Maschinen übertragen lässt und Cyborgs, Mensch-Maschinen, dann die nächste Stufe der Evolution bilden.



Die Unterscheidung zwischen starker und schwacher KI hat der US-amerikanische Sprachphilosoph John Robert Searle vor rund vier Jahrzehnten vorgeschlagen. Starke KI ist bis auf Weiteres Science Fiction. Schwache Künstliche Intelligenz hingegen ist im Hier und Jetzt am Werk, wenn ein Computersystem Aufgaben übernimmt, von denen man noch bis vor Kurzem dachte: Das kann nur ein Mensch erledigen, der in irgendeiner Form auch seinen Kopf anstrengen muss. Oft geht es dabei um althergebrachte Aufgaben der Wissensarbeit, Fallbearbeitung in Versicherungen zum Beispiel oder das Schreiben von Nachrichten- oder Sportmeldungen. Eingebettet in physische Maschinen macht KI nicht nur Autos, sondern Fabriken, landwirtschaftliches Gerät, Drohnen oder Rettungs- und Pflegeroboter funktionstüchtiger.



Es gibt immer wieder Parallelen zu menschlichem Verhalten, aber zu einem pragmatischen Verständnis von KI gehört, dass die schlaunen Maschinen die Vorgehensweise des Menschen beim Lösen der Aufgaben nicht imitieren müssen – und schon gar nicht die biochemischen Vorgänge im menschlichen Gehirn. Sie verfügen in der Regel über die Fähigkeit, eigenständig mathematische Lösungswege zu erkunden, ihnen dafür vorgegebene Algorithmen zu verbessern oder gar eigenständig Algorithmen zu entwickeln. Das Ergebnis ist dann, dass die Maschine den Job besser, schneller oder günstiger erledigt als der Mensch. Je größer wiederum die Überlegenheit der Maschine über menschliche Problemlöser ist, desto schneller verbreiten sich die Systeme. Dies geschieht allerdings nicht zu Null-Grenzkosten nach dem Prinzip, eine digitale Kopie kostet nichts, wie die Apostel der digitalen Revolution behaupten. Digitale Technologie ist teuer, und das wird auch noch eine Weile so bleiben. Aber empirisch nachweisbar ist: Die Zyklen der Einführung und Verbreitung neuer Technologien werden kürzer.



Die für die Menschheit noch drängendere Frage könnte freilich sein: Wird sich eine starke KI, eine Super-Intelligenz entwickeln, die sich autonom und in Feedbackschleifen ein immer besseres Bild von der Welt und sich selbst errechnet? Ein System, »das den Menschen von der Spitze der Denkerkette vertreibt«, wie Nick Bostrom formuliert, der Leiter des Future of Humanity Institute in Oxford. Die Folge wäre: Das superintelligente System wäre vom Menschen nicht mehr kontrollierbar. Und könnte diese Superintelligenz sich gar wie im Science Fiction gegen den Menschen wenden und am Ende die Maschine den Menschen auslöschen?



Die gute Nachricht gleich vorab: Künstlich intelligente Systeme werden den Menschen vorerst nicht unterjochen. Der Weltuntergang fällt auch diesmal aus. Aber KI-Systeme haben eingebaute Schwächen, die sie anfällig für Fehlentscheidungen machen, ihrem Einsatz Grenzen setzen und uns Menschen in die Pflicht nehmen, ihr algorithmisches Wirken stets kritisch zu hinterfragen. Verblüffend dabei ist, wie menschlich die Schwächen von KI-Systemen wirken. So tendieren auch neuronale Netze zu Vorurteilen, die nicht vom Entwickler einprogrammiert wurden, sondern sich implizit aus den Trainingsdaten ergeben können. Wenn zum Beispiel eine KI-gestützte Kreditvergabe aufgrund der Trainingsdaten zu erkennen meint, dass eine ethnische Minderheit oder Männer über 53,8 Jahre oder Radfahrer mit gelben Helmen und 8-Gang-Schaltung Kredite weniger zuverlässig zurückzahlen, wird sie es bei ihrem Scoring-Modell berücksichtigen – egal ob dies illegal ist oder vollkommen unsinnig. Das erlernte Vorurteil ist umso gefährlicher, weil die Maschine es nicht offenlegt.



Die gute Nachricht gleich vorab: Künstlich intelligente Systeme werden den Menschen vorerst nicht unterjochen. Der Weltuntergang fällt auch diesmal aus. Aber KI-Systeme haben eingebaute Schwächen, die sie anfällig für Fehlentscheidungen machen, ihrem Einsatz Grenzen setzen und uns Menschen in die Pflicht nehmen, ihr algorithmisches Wirken stets kritisch zu hinterfragen. Verblüffend dabei ist, wie menschlich die Schwächen von KI-Systemen wirken. So tendieren auch neuronale Netze zu Vorurteilen, die nicht vom Entwickler einprogrammiert wurden, sondern sich implizit aus den Trainingsdaten ergeben können. Wenn zum Beispiel eine KI-gestützte Kreditvergabe aufgrund der Trainingsdaten zu erkennen meint, dass eine ethnische Minderheit oder Männer über 53,8 Jahre oder Radfahrer mit gelben Helmen und 8-Gang-Schaltung Kredite weniger zuverlässig zurückzahlen, wird sie es bei ihrem Scoring-Modell berücksichtigen – egal ob dies illegal ist oder vollkommen unsinnig. Das erlernte Vorurteil ist umso gefährlicher, weil die Maschine es nicht offenlegt.



Julian Nida-Rümelin
Nathalie Weidenfeld

DIGITALER HUMANISMUS

Eine Ethik für das Zeitalter der
Künstlichen Intelligenz





Warum KIs nicht denken können

Computer und Menschen denken nicht auf die gleiche Weise. Oder, um es zu präzisieren : Ein Computer denkt in unserem Sinne überhaupt nicht. Im Grunde müsste schon angesichts auffälliger Unterschiede zwischen künstlicher und menschlicher Intelligenz klar sein, dass Computer zwar Denken erfolgreich simulieren, ja viele menschliche Denkvorgänge, zum Beispiel algebraische Operationen, weit präziser und schneller vornehmen können als Menschen (das beginnt schon beim Taschenrechner). Aber trotz dieser oft perfekten Simulation liegt beim Computer kein eigenes verständiges Erfassen, kein Problembewusstsein, keine Einsicht zugrunde.



Wenn Internetanbieter eine Bestätigung haben wollen, dass der Nutzer kein Computer ist, fragen sie zum Beispiel danach, auf welchem der folgenden Abbildungen ein Straßenschild zu sehen ist oder ein Auto oder ein Haus. Diese schlichten, ja idiotensicheren Fragen kann jedes Kind sofort und zuverlässig beantworten. Da visuelle Softwareprogramme Erkenntnisvorgänge dieser Art nur simulieren, aber nicht selbst über eine eigene Wahrnehmungsfähigkeit verfügen, versagen diese schon angesichts solch schlichter Aufgaben. Ähn- / Reakliches gilt für das Ärgernis der digitalen Übersetzungsprogramme. An ihnen wird nun schon seit Jahrzehnten intensiv gearbeitet, Linguistik und Mathematik verbindet hier ein gigantisches Forschungs- und Entwicklungsprogramm, und doch sind die Ergebnisse anhaltend kläglich. Es fehlt eben an Sprachverständnis. Eine Software versteht keinen eigenen Satz, selbst wenn es ihr einmal gelingen sollte, ihn korrekt zu übersetzen.



Nun stellt sich allerdings die Frage, was eigentlich den kategorialen Unterschied ausmacht zwischen der bloßen Anwendung algorithmengesteuerter Verfahren, etwa bei visueller Erkennungssoftware oder Übersetzungsprogrammen, und dem Erfassen von Sinn und Bedeutung. Um zu erklären, was hiermit gemeint ist, hilft ein Ausflug in die Mathematik und Logik der 30er-Jahre. In dieser Zeit entwickelte Kurt Gödel ein Theorem, das bis heute als das wichtigste Ergebnis der formalen Logik und Meta-Mathematik gilt.



Kurt Gödel (1906-1978)



Dieses Theorem besagt, dass es wahre logische und mathematische Sätze gibt, die in dem Sinne nicht berechenbar sind, als es kein algorithmisches Verfahren gibt, das es erlaubt, die Richtigkeit dieser Sätze zu beweisen. Damit ist die Annahme, dass es einen Algorithmus geben könnte, der das (menschliche) Denken als Ganzes repräsentieren könnte, als falsch erwiesen. Das heißt keineswegs, dass es nicht möglich sei, die Richtigkeit oder Falschheit von Hypothesen und Überzeugungen zu überprüfen, es heißt lediglich, dass es keinen Algorithmus gibt, der uns diese Überprüfung abnimmt: Wir müssen schon selber denken. Wir können nur diejenigen Ausschnitte unserer Entscheidungspraxis an Computer oder von Computern gesteuerte / Roboter delegieren, die sich durch Algorithmen nachbilden lassen.



Das Theorem der Unberechenbarkeit von Kurt Gödel zeigt uns, dass die Welt der logischen und mathematischen Strukturen insgesamt nicht algorithmisch strukturiert ist. Menschliche Vernunft, die menschliche Fähigkeit, Überzeugungen, Entscheidungen und emotive Einstellungen zu begründen und auf dieser Grundlage ein kohärentes Weltbild und eine kohärente Praxis zu entwickeln, lässt sich nicht im Modell eines digitalen Computers erfassen. Es wird nie gelingen, die hohe Komplexität unserer lebensweltlichen Begründung vollständig und in adäquater Weise formal zu erfassen. Roboter und Softwaresysteme funktionieren nach einem Algorithmus, Menschen nicht. Darin liegt einer ihrer zentralen Unterschiede begründet.



Wir müssen uns klarmachen, dass das » Denken «, das » Rechnen «, die » Reaktionen «, die » Entscheidungen « eines Roboters nur Simulationen von Denken, Rechnen, Reaktionen, Entscheidungen sind und keine – im menschlichen Sinne – echten Denkvorgänge. Nehmen wir das Beispiel des Schachcomputers. Es gibt wenig Ähnlichkeit zwischen dem Denken eines Menschen und dem » Denken « eines Schachcomputers. Wären die » Denkvorgänge « ähnlich oder gar gleich, hätte nie ein menschlicher Schachspieler auch nur eine minimale Chance gegen einen Computer. Das menschliche Gehirn ist nämlich restlos überfordert, wenn es auch nur einen winzigen Teil der möglichen Stellungen durchrechnen müsste, die auch die schlichtesten Schachcomputer durchrechnen. Tatsächlich aber spielt das Durchrechnen aller möglichen Nachfolgekonstellationen und die der möglichen späteren eigenen Reaktionen auf dem Schachbrett nach einem bestimmten Zug für menschliche Schachspieler keine Rolle. Sie beschränken sich auf einige wenige relevante Optionen und können im Gegensatz zum Schachcomputer nur wenige Züge im Voraus kalkulieren. Der von den Schachregeln definierte Möglichkeitsraum von Nachfolgekonstellationen auf dem Schachbrett ist so gigantisch, dass auch der intelligenteste Schachspieler diesen nicht im Ansatz überblicken kann.



Doch auch wenn die neuesten Schachcomputer praktisch unbesiegbar sind, darf dies nicht als Beleg dafür genommen werden, dass Roboter das Gleiche leisten wie menschliche Gehirne. Roboter sind so konstruiert, dass sie menschliches Denken in den Begriffen der Computersprache (Software, Hardware, neuronale Netze, binäre Logik etc.) simulieren, sie haben selbst keine mentalen Eigenschaften, sie erfassen keine Figurenkonstellation auf dem Schachbrett, wie ihr gelegentliches Versagen in simplen Konstellationen schlaglichtartig verdeutlicht.



Doch was ist, wenn Roboter immer komplexer und fortschrittlicher werden ? Wie etwa die vom Deep- Mind-Forschungszentrum der Firma Google entwickelte Künstliche Intelligenz namens AlphaGo, die darauf programmiert wurde, das chinesische Brettspiel Go perfekt zu beherrschen ? Aufgrund der Vielzahl der möglichen Positionen stellt Go im Vergleich zu Schach eine wesentlich größere Herausforderung für Programmierer dar. Während ein Schachspieler bei jedem Zug ungefähr 35 Handlungen vollziehen kann, sind es bei Go 250. Ein weiterer Unterschied : Ein durchschnittliches Schachspiel dauert 80 Züge, bei einem Go-Spiel sind es 150. Doch 2016 war die Sensation perfekt : AlphaGo besiegte Lee Sedol, den wahrscheinlich besten Go-Spieler der Welt.



Bedeutet der Übergang von Softwaresystemen, die ihre Leistung durch das Ausrechnen einer gewaltigen Vielfalt möglicher Konstellationen beziehen, zu Systemen, die » selbst lernen «, aufgrund vorgegebener Regeln eigene Regeln zu entwickeln, dass Künstliche Intelligenz ab diesem Zeitpunkt menschliches Denken nicht nur simuliert, sondern auch selbst als genuines Denken interpretiert werden sollte ?



Es gibt in der Tat die verbreitete Auffassung, dass mit der Einführung der künstlichen neuronalen Netze in der Computertechnologie die Festschreibung von Computern als Turing-Maschinen überwunden sei. Das ist jedoch ein Irrtum. Sowohl die Top-down- Methode der Rechenprozesse (» Computation «) als auch die Bottom-up- Methoden selbstlernender Systeme sind von Algorithmen geleitet. Auch bei selbstlernenden Systemen muss vorab feststehen, welche Resultate erwünscht sind, um den » Lernprozess « des Computers in Gang zu setzen. Ziel ist es, die gewünschten Ergebnisse aufgrund bestimmter Eingabedaten zu erreichen. Ein Beispiel sind die schon recht weit gediehenen Softwaresysteme zur Gesichtserkennung.



Der Ausdruck » neuronale Netze « ist dabei in doppelter Hinsicht irreführend. Erstens bestehen diese Netze nicht aus Neuronen, sondern aus Leiterstrukturen, und zweitens ähneln diese Netze allenfalls sehr entfernt der immensen Komplexität und Plastizität des menschlichen Gehirns. Da das Funktionieren der (realen) neuronalen Netze des menschlichen Gehirns noch ganz unzureichend erfasst ist, kann auch nicht die Rede davon sein, dass die Computertechnik menschliche Denkvorgänge beziehungsweise deren neuronale Realisierung imitiert.



Das gilt auch für das sogenannte Deep Learning. Damit wird eine Lernmethode bezeichnet, mit der Softwaresysteme dadurch aus Erfahrung lernen können, dass sie eine Reihe hierarchisch strukturierter Konzepte nutzen. Die Informationen werden von dem System von einer Schicht an die nächste Schicht weitergegeben und dabei selbstständig verarbeitet. Dabei werden die Merkmale zunehmend abstrakt, und das System muss selbst » entscheiden «, welche Konzepte für die Erklärung nützlich sind. Die hohe Komplexität dieses Systems ändert nichts an ihrem algorithmischen Charakter, bemerkenswert ist dagegen ein anderer / Aspekt : Mit der steigenden Komplexität ist ein massiver Verlust an Transparenz verbunden, mit der Folge, dass für den menschlichen Beobachter, auch für den Programmierer, nicht mehr nachvollziehbar ist, auf welchem Weg der Lernprozess erfolgreich war, welche Regeln sich das System aufgrund vorgegebener Meta- Regeln oder Meta-Meta-Regeln gegeben hat. Im Extremfall würde es zu einer Blackbox, deren Output bei gegebenem Input bekannt ist, die Regeln der Korrelation aber nicht.



Auch wenn die Bottom-down-Computer oft um viele Zehnerpotenzen bessere Resultate erzielen als die entsprechenden menschlichen Denkprozesse (etwa für arithmetische Operationen generell oder für die Berechnung von Funktionsgleichungen oder geometrischer Figuren), liegen gerade die künstlichen neuronale Strukturen simulierenden Netze in der Regel weit unter den menschlichen Fähigkeiten : Menschen sind noch immer weit besser im Gesichtserkennen als selbst die am weitesten entwickelten Softwaresysteme, und der Gang humanoider Roboter ist auch nach langwierigen » Selbstlernprozessen « weit weniger elegant und vielfältig als der des Menschen.



Das letzte, aber womöglich wichtigste Argument gegen den Versuch, einer Rechenmaschine menschliches Denken zuzuschreiben, ist folgendes : Wenn wir Menschen einen Denkvorgang beziehungsweise theoretische wie praktische Intelligenz zuschreiben, dann berücksichtigen wir nicht nur eine Vielfalt von mentalen Eigenschaften, sondern auch Intentionen, also das Gerichtetsein des Geistes auf etwas. Diese Intentionalität aber wird durch künstliche neuronale Netze nicht realisiert.



Der amerikanische Philosoph John Searle hat in Bezug auf diese Frage ein berühmtes Gedankenexperiment namens » Das chinesische Zimmer « entwickelt. In diesem Gedankenexperiment geht es darum, sich einen Menschen vorzustellen, der in einem geschlossenen Raum sitzt und kein Chinesisch spricht, nicht einmal die Schriftzeichen kennt. Dieser Person werden nun durch den Türschlitz Papierschnipsel mit chinesischen Zeichen zugeführt, auf die sie ebenfalls mit chinesischen Schriftzeichen antworten muss. Dazu erhält sie eine Art Anleitung für die richtigen Antwortzeichen sowie ein Handbuch in ihrer Muttersprache, das es ihr erlaubt, anhand der zugeführten Symbole eine Antwort auf Chinesisch zu schreiben. Sie folgt hierbei aber ausschließlich den Anweisungen aus der Anleitung und versteht folglich die Antworten, die sie anschließend durch den Türschlitz wieder nach draußen schiebt, nicht. Draußen vor dem Schlitz steht ein chinesischer Muttersprachler, der, nachdem er die Fragen formuliert, auf die Papierschnipsel geschrieben und anschließend die Antworten darauf bekommen hat, zu dem Schluss kommt, im Raum müsse sich jemand befinden, der ebenfalls Chinesisch spricht.



John Searle (* 1932)



„Das Entscheidende, was hier fehlt, ist offenkundig : Es ist das Verständnis der chinesischen Sprache. Auch wenn ein System – hier das » chinesische Zimmer « – funktional äquivalent ist mit jemandem, der Chinesisch versteht, versteht dieses System noch nicht Chinesisch. Chinesisch verstehen und sprechen setzt eine Vielzahl von Kenntnissen voraus. Eine Person, die Chinesisch spricht, bezieht sich mit bestimmten Ausdrücken auf die entsprechenden Gegenstände. Sie verfolgt mit bestimmten Äußerungen bestimmte, dazu pas- / sende Absichten. Sie bildet aufgrund von (in chinesischer Sprache) Gehörtem bestimmte Erwartungen aus etc. All diese Eigenschaften hat das chinesische Zimmer nicht. Es verfolgt keine Absichten, und es hat keine Erwartungen, die belegen, dass es die chinesische Sprache spricht und versteht. Mit anderen Worten : Das chinesische Zimmer simuliert das Verständnis des Chinesischen, ohne selbst der chinesischen Sprache mächtig zu sein



Searle hat dieses Argument Jahre später radikalisiert. 43 In diesem zweiten Argument verbindet er seinen philosophischen Realismus, also die These, dass es eine Welt gibt, die unabhängig davon existiert, ob sie beobachtet wird, mit einer sogenannten » intentionalistischen Zeichentheorie «. Diese besagt, dass Zeichen immer nur Bedeutung für uns Menschen haben, die wir die Zeichen nutzen und interpretieren. Wir tun dies dadurch, dass wir uns darauf verständigen und darauf einigen, dass diese Buchstaben oder Symbole für etwas stehen. Ohne diese konventionellen Setzungen oder etablierten Praktiken des Gebrauchs der Zeichen haben diese keine Bedeutung. Insofern ist es irreführend, den Computer als eine zeichenverarbeitende oder syntaktische Maschine zu verstehen, die bestimmten logischen oder grammatischen Regeln folgt.



Der Computer einigt sich nicht mit anderen Computern oder Menschen auf Bedeutungen, er folgt vorgegebenen Regeln. Ein Computer besteht lediglich aus unterschiedlichen, physikalisch beschreibbaren Elementen, von denen einige Strom leiten und andere nicht. Die Rechenprozesse sind eine Abfolge von elektrodynamischen und elektrostatischen Zuständen. Diesen Zuständen werden dann Zeichen zugeordnet, die wir mit / bestimmten Interpretationen und Regeln unterlegen. Die physikalischen Prozesse im Computer haben keine Syntax, sie » kennen « keine logischen oder grammatischen Regeln, sind keine Zeichenfolgen. Insofern ist die syntaktische Interpretation beobachterrelativ. Wir als Computernutzer und -programmierer gestalten die elektrodynamischen Prozesse so, dass sie für uns einer Syntax entsprechen.



Dieses Argument ist radikal, einfach und zutreffend. Es beruht auf einer realistischen Philosophie und einer mechanistischen Interpretation der Computer. Es bricht mit der verbreiteten gemeinsamen Auffassung bei Anhängern der KI und ihren Gegnern, dass Computer syntaktische Maschinen sind. Computer sind das, was sie materiell sind : Gegenstände, die sich mit den Mitteln der Physik vollständig beschreiben und erklären lassen. Syntax ist kein Teil der Physik, die Physik beschreibt keine Zeichen, keine grammatischen Regeln, keine logischen Schlüsse, keine Algorithmen.



Dieses Argument ist radikal, einfach und zutreffend. Es beruht auf einer realistischen Philosophie und einer mechanistischen Interpretation der Computer. Es bricht mit der verbreiteten gemeinsamen Auffassung bei Anhängern der KI und ihren Gegnern, dass Computer syntaktische Maschinen sind. Computer sind das, was sie materiell sind : Gegenstände, die sich mit den Mitteln der Physik vollständig beschreiben und erklären lassen. Syntax ist kein Teil der Physik, die Physik beschreibt keine Zeichen, keine grammatischen Regeln, keine logischen Schlüsse, keine Algorithmen.



Alles Gute und bis zur nächsten Sitzung!