

## Die Macht der Vermenschlichung und die Ohnmacht der Begriffe

In den siebziger Jahren des letzten Jahrhunderts formulierte der britische Physiker und Science-Fiction-Schriftsteller Arthur C. Clarke, dass jede hinreichend fortschrittliche Technologie von Magie nicht zu unterscheiden sei. Dabei bezog er sich zwar auf technische Konzepte wie Antigravitationsantriebe, „fließende Straßen“ oder Atombaumaschinen, bei denen die Menschen in manchen seiner Geschichten selbst nicht mehr wussten, wie und warum sie funktionierten. Doch in etwas abgewandelter Form könnte man diese Aussage auch auf andere Art kritisch verwenden: Jede hinreichend oft mit menschlichen Begriffen beschriebene Technik wird bald auch für menschenähnlich gehalten. Das ist nicht nur philosophisch hoch problematisch.

Seit einigen Jahren erleben die künstlichen neuronalen Netze eine wahre Renaissance. Obwohl die theoretischen Grundlagen der aktuell verwendeten Konzepte schon in den 1980er-Jahren erdacht wurden, zeitigen die neuerlichen technischen Rahmenbedingungen und die stetig anwachsende Datenmenge nun vielfach praktisch nutzbare Erfolge. Von der Spracherkennung über die Satzanalyse bis hin zur Bildinhaltsbestimmung. Insbesondere auf riesigen Datenbeständen kann man sie gut anwenden.

### Grundaufbau künstlicher neuronaler Netze

Künstliche neuronale Netze sind ein Ansatz der Informatik, um komplexe Probleme zu lösen. Der Aufbau wurde von der Funktionsweise des Gehirns inspiriert, allerdings ist er sehr reduziert modelliert. Die kleinsten Einheiten, die „Neuronen“, haben jeweils mehrere Signaleingänge und einen Signalausgang. In jedem Neuron werden die Signale der Eingänge gewichtet, aufsummiert und dann am Ausgang ausgegeben. Diese Ein- und Ausgänge der Neuronen sind im Allgemeinen in Reihen („Schichten“) miteinander verbunden, sodass ein Netzwerk entsteht. Dieses künstliche neuronale Netz hat also als Input die Eingänge der ersten Schicht und als Output die Ausgänge der letzten Schicht, dazwischen liegt eine Anzahl „versteckter“ Schichten (Abbildung 1).

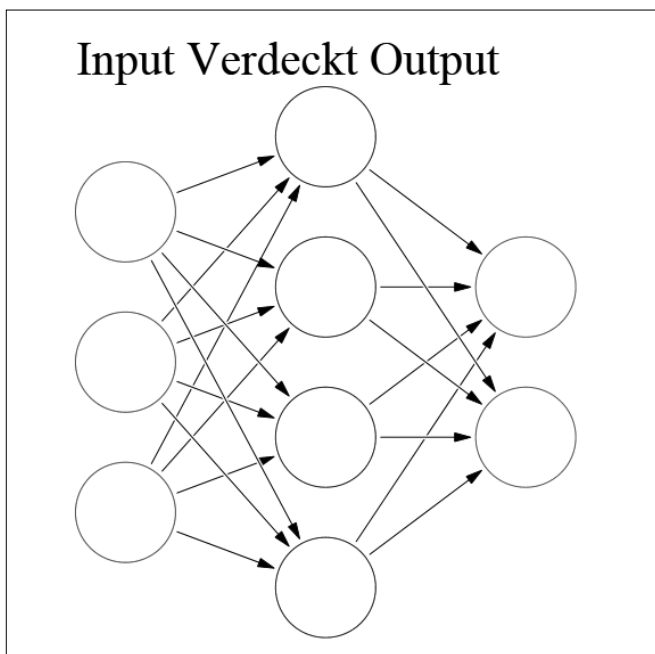


Abbildung 1: Künstliches neuronales Netz, schematisch, CC-BY-SA-3.0, nach Glosser.ca [https://en.wikipedia.org/wiki/Artificial\\_neural\\_network#/media/File:Colored\\_neural\\_network.svg](https://en.wikipedia.org/wiki/Artificial_neural_network#/media/File:Colored_neural_network.svg)

### Training der Netze

Soll nun ein solches Netz z. B. Katzen auf Bildern finden, gibt man die Pixel eines Bildes als Input ein und ändert die Gewichtungen der Neuronen ein wenig, bis am Ausgang das gewünschte Signal anliegt; also eine Katze signalisiert wird (Stichwort *Backpropagation*). Dies wird so oft mit sehr vielen Katzenbildern gemacht, bis sich die Gewichtungen (fast) nicht mehr ändern. Im besten Falle sind das Bilder von Katzen aus ganz verschiedenen Perspektiven in verschiedenen Sichtverhältnissen usw., sodass sich „die Katze an sich“ im Netzwerk niederschlägt und nun neue, unbekannte Bilder besser analysiert werden können (ähnlich funktioniert es mit Spracherkennung und dergleichen). Nach dieser Kalibrierungsphase hat man das Netz also so trainiert, dass es Katzenbilder ... nun ja, was eigentlich? Was wäre hier der passende Begriff?

### Begriffsdomänen und Alltagssprache

In der Tat werden an dieser Stelle in den Computerwissenschaften Begriffe wie „Erkennen“, „Lernen“ oder sogar „Verstehen“ verwendet. Das ist zwar grundsätzlich unproblematisch, denn viele Wissenschaften verwenden Fachbegriffe, die von alltäglichen Wörtern abgeleitet sind. In den jeweiligen Wissenschaften haben diese Worte dann aber klar definierte Bedeutungen oder zumindest fest damit verknüpfte Diskurse. Beispiele sind die Begriffe Angst (Psychologie), Impuls (Physik), Wille (Philosophie) oder auch Verwerfung (Geologie) und Reihe (Mathematik). Oft haben die gleichen Worte in verschiedenen Domänen ganz unterschiedliche, teilweise sogar gegensätzliche Bedeutungen, wie die Beispiele Arbeit (Physik und Wirtschaftstheorie) und Transparenz (Informatik und Politikwissenschaften) verdeutlichen.

Problematisch wird es jedoch erst, wenn diese wissenschaftlichen Begriffe unreflektiert in andere Domänen oder zurück in die Alltagssprache projiziert werden, sei es durch schlechten Wissenschaftsjournalismus, absichtliche Ungenauigkeit zum Zwecke der Drittmittelinwerbung, im Rahmen von PR-Maßnahmen oder allgemein aufgrund fehlender Sensibilität bezüglich der verschiedenen Bezugsebenen.

Auf unser Beispiel bezogen zeigt sich also, wie vielschichtig die Aussage „Das neuronale Netz hat gelernt, Katzenbilder zu erkennen“ verstanden werden kann.

### „Inception“

Betrachten wir nun die Beschreibungen um Googles *Deep Dream*-Bildererkennungs- und -klassifikationssoftware, Codename *Inception*.

Bislang ist immer noch größtenteils unbekannt, wie und warum gut trainierte künstliche neuronale Netze wirklich funktionieren. Im Gegensatz zur traditionellen Modellierung von Sachverhalten und der anschließenden Repräsentation in wohldefinierten Datenstrukturen haben künstliche neuronale Netze die antrainierten Eigenschaften irgendwie in ihrer Struktur verteilt gespeichert.

Um zu verstehen, wie das funktioniert, wurde Netzen, die sonst Bilder klassifizieren, buntes Rauschen vorgesetzt (Abbildung 2). Da die Netze die dortigen Strukturen analysierten und daraus Bilder konstruierten, waren die Ergebnisse sehr psychedelisch.<sup>1</sup>

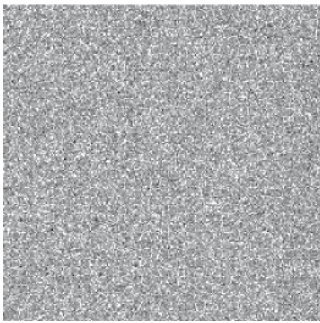


Abbildung 2:  
Buntes Rauschen

Interessant hierbei sind die Beschreibungen. Allein der Name *Deep Dream* ist bezeichnend, aber auch *Inside an artificial brain* und *Inceptionism* lassen der Fantasie freien Lauf. In den Texten dazu häufen sich Formulierungen wie das „Netz trifft Entscheidungen“, es „sucht“ nach den richtigen Eigenschaften in Bildern, es „lernt wie ein Kind“ oder sogar, dass es Bilder „interpretiert“. Auch wenn diese Netze allein irgendwelche strukturellen Eigenschaften von Bildern inferieren, erhofft man sich zusätzlich, etwas über die Grundlagen menschlicher Kreativität zu lernen.<sup>2</sup> Um die konkrete Behandlung der durchaus interessanten Fragen, warum und inwieweit diese Beschreibungen unzutreffend sind, soll es hier jedoch weniger gehen. Thema dieses Textes sind eher die dadurch entstehenden Konsequenzen, denn werden (Fach-)Begriffe so undifferenziert verwendet und vermischt, scheint der entstandene „Schaden“ zunächst rein akademisch, doch diese Sicht greift zu kurz.

### Zuschreibungsmechanismen

Informationstechnische Systeme werden naturgemäß in den seltensten Fällen wegen ihrer tatsächlichen Eigenschaften eingesetzt, sondern aufgrund der vermuteten Funktion. Das ist wegen der Komplexität dieser Systeme wenig verwunderlich und lässt sich gut am rasant wachsenden „IT-Sicherheits“-Markt ab-



**Rainer Rehak** beschäftigt sich seit ca. zehn Jahren mit dem Themenfeld *Informatik und Gesellschaft*. Er studierte Informatik und Philosophie in Berlin, Hong Kong und Peking. Während des Studiums arbeitete er am Lehrstuhl für *Informatik in Bildung und Gesellschaft* von Wolfgang Coy. Aktuell lehrt er an der HTW Berlin und der HS Harz in den Bereichen Datenschutz und Datensicherheit, Informatik und Gesellschaft sowie Netzwerke. In der echten Praxis ist er als IT-Sicherheitsberater und Unix-Serveradministrator tätig. Er ist Träger eines FIFF-Studienpreises.

<https://gewissensbits.gi.de/rainer-rehak/>

lesen: Größtenteils nutzlose Produkte werden genauso angeboten und gekauft, wie tatsächlich sinnvolle, funktionierende Software. Den Unterschied erkennen – wenn überhaupt – nur Expertinnen und Experten der Branche.

Natürlich sind auch rechtlich und politisch relevante Entscheidungen von technischen Vorstellungen abhängig. Ein Beispiel: Eine staatlich genutzte Trojanersoftware (heimliche Online-durchsuchung und Quellen-TKÜ) darf rechtlich nur dann zur Anwendung kommen, wenn sie kontrollierbar ist und auf bestimmte Funktionen beschränkt werden kann. Der behördliche Einsatz hängt nun nicht davon ab, ob die Software tatsächlich die geforderten Eigenschaften besitzt, sondern ob ihr die Eigenschaften zugeschrieben werden. In der 2011 durch den Chaos Computer Club (CCC) bekannt gemachten DigiTask-Trojanersoftware wurden zahlreiche illegale Funktionalitäten aufgedeckt, von denen einige prinzipieller Natur waren. Diese prinzipiellen Probleme hätten den Einsatz solcher Produkte grundsätzlich verhindern müssen, aber die zuständigen Personen hatten der Software andere, ausschließlich rechtskonforme Eigenschaften zugeschrieben.

### Die Anthropomorphie der Zahlen

Verbinden wir diese Erkenntnisse nun mit der Vermenschlichung von künstlichen neuronalen Netzen, so ergibt sich durch die beschriebenen Zuschreibungsmechanismen ein differenziertes Bild der Konsequenzen nachlässiger Begriffsverwendung. Sollten relevante Personen in Politik und Gesellschaft wirklich irgendwann davon überzeugt sein, dass diese „neuartigen“ künstlichen neuronalen Netze ein Verständnis von Dingen entwickeln oder Sachverhalte ordentlich interpretieren können, stünde ja auch der Nutzung für hoheitliche Aufgaben nichts im Wege. In Diskussionen um die Automatisierung von Hartz-IV-Entscheidungen oder anderer rechtlicher Abläufe werden sich dann auch Informatik-treibende aus diesen Themenfeldern einmischen müssen.

Man wird sehen, ob wir die Netze bis dahin weit besser verstanden haben werden oder wenn nicht, inwieweit ihre Funktionsweise von Magie zu unterscheiden ist.



### Anmerkungen

- <https://www.youtube.com/watch?v=SCE-QeDfXtA>
- <http://googlresearch.blogspot.de/2015/06/inceptionism-going-deeper-into-neural.html>

**Rainer Rehak**