

KI – der große Etikettenschwindel

Mythos und Realität semiotischer Maschinen

Peter Brödner

Interdisziplinäres Gespräch »Kreativität und Technik«
Universität Leipzig | 3.07.2018

Ungewisser Gegenstand (I): Was ist ein »KI-System«?

Übliche KI-Definition:

Computing systems »that are capable of **performing tasks** commonly **thought to require intelligence**. **Machine learning** ... refers to the development of digital systems that **improve** their **performance** on a given task over time **through experience**« (Autorengruppe 2018).

Niemand weiß, was Intelligenz genau ist. Als System zur »**Maschinisierung von Kopfarbeit**« (Nake 1992) bewältigt doch **jeder Computer** Intelligenz erfordernde Aufgaben, mithin lassen sich »KI-Systeme« **nicht** von anderen Computersystemen **unterscheiden**. Zudem ist **Lernen** aus **Erfahrung** lebendiger Organismen **nur bedingt formalisierbar** und durch tote berechenbare Funktionen nachbildbar, allenfalls lassen sich **adaptive Systeme** realisieren, die sich mittels Sensoren algorithmisch gesteuert ihrer **Umwelt anpassen**.

Um weitere ähnliche Definitionen steht es nicht besser:

AI research is defined as the study of »**intelligent agents**: any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.« Colloquially, the term »artificial intelligence« is applied when a machine **mimics »cognitive« functions** that humans **associate** with other human minds, such as »learning« and »problem solving« (Russell & Norvig 2009).

Zwar kann **mangels** hinreichender **Unterscheidbarkeit** niemand wissen, **was ein »KI-System«** eigentlich ist, gleichwohl werden für dieses **Phantom** präzise **Umsatzprognosen** erstellt:

»Global business value derived from artificial intelligence (AI) is projected to total \$1.2 trillion in 2018, an increase of 70 percent from 2017, according to Gartner, Inc. AI-derived business value is forecast to reach \$3.9 trillion in 2022« (Gartner 25.04.2018).

Ungewisser Gegenstand (2): Was ist ein ›KI-System‹?

»The *artificial intelligence* problem is taken to be that of *making a machine behave in ways that would be called intelligent if a human were so behaving.*« – so bereits **McCarthy** 1955 in seinem Antrag an die Rockefeller Foundation für das **Dartmouth Summer Research Project on AI**.

Dieser Wahnsinn ist Methode des ›**computational model of the mind**‹:

Der sog. ›intentionalen Einstellung‹ (›*intentional stance*‹; Dennett 1987) zufolge werden ›**intelligenten Agenten**‹ **intentionale Zustände** wie Überzeugungen, Wünsche oder Absichten **zugeschrieben**, ausdrücklich um deren **undurchschaubares Verhalten** einfacher und **verständlicher beschreiben** zu können:

»To ascribe beliefs, free will, intentions, consciousness, abilities, or wants to a machine is **legitimate** when such an **ascription expresses the same information about the machine that it expresses about a person**. It is useful when the ascription helps us understand the structure of the machine, its past or future behaviour, or how to repair or improve it. It is perhaps never logically required even for humans, but expressing reasonably briefly what is actually known about the state of the machine in a particular situation may require mental qualities or qualities isomorphic to them.« (McCarthy 1978, zit. nach Shoham 1993: 53)

Der propagandistische **Trick**:

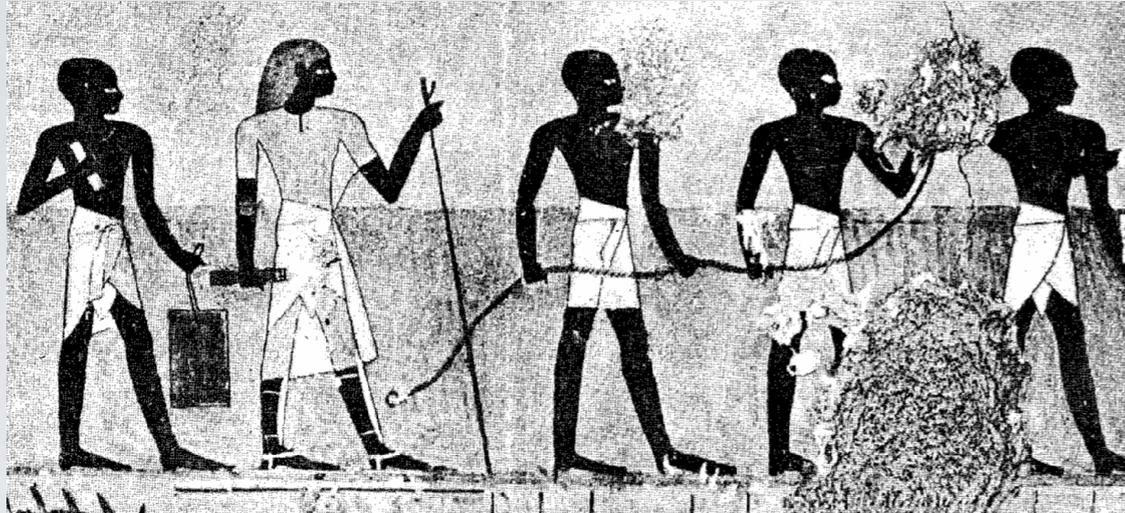
Was als **Eigenschaft** maschinellen Verhaltens (›intelligent‹, ›intentional‹) durch dieses selbst erst zu erzeugen wäre, wird diesem lediglich **ein-** und **zugeschrieben**, aber **nicht von selbst** hervorgebracht. **Intelligenz** zeigt sich allein in der **Entwicklung** von **Methoden** der **Problemlösung** in Form berechenbarer Funktionen, die dann Computer lediglich ausführen (vgl. vertikale Arbeitsteilung).

»Al is whatever hasn't been done yet.«

(Hofstadter 1979: 601)

Kulturgeschichtliches Beispiel: ›Pythagoras‹

Die Aufgabe: Erzeugung eines genauen rechten Winkels in der freien Natur (etwa um Felder abzu- stecken und zu vermessen). Lösung mittels **pythagoräischer Zahlentripel** und einem entsprechend eingeteilten **Seil**:



Ägyptische Landvermesser um 1475 v. u. Z.

Die Aufgabe:

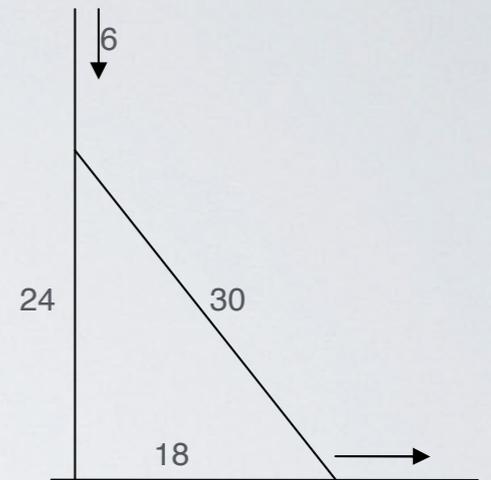
- (1) Ein Balken 30.
- (2) Von oben ist er 6 herabgekommen.
- (3) Von unten was hat er sich entfernt.

Die Lösung:

- (4) 30 quadriere. 900 siehst du.
- (5) 6 von 30 abgezogen. 24 siehst du.
- (6) 24 quadriere. 576 siehst du.
- (7) 576 von 900 ziehe ab. 324 siehst du.
- (8) 324 hat was als Quadratwurzel: 18.

Das Ergebnis:

- (9) 18 am Boden hat er sich entfernt.



$$a^2 + b^2 = c^2$$

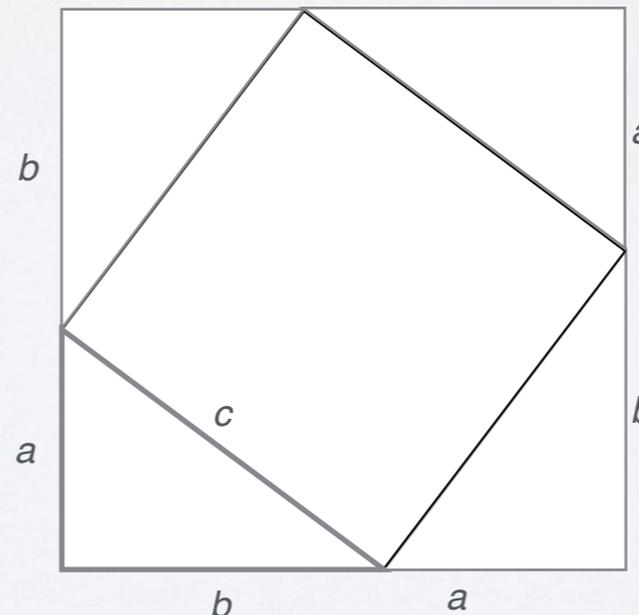
(Quelle: British Museum. Deutsche Übersetzung einer > 4000 Jahre alten babylonischen Keilschrifttafel; Zahlenangaben ins Dezimalsystem übertragen.)

Beweis durch Kombination von anschaulichem und begrifflichem Denken: Die große Quadratfläche beträgt $(a + b)^2$; sie setzt sich zusammen aus der Fläche c^2 und vier rechtwinkligen Dreiecken mit jeweils dem Flächeninhalt $(a b)/2$, folglich gilt:

$$(a + b)^2 = c^2 + 4 (a b)/2$$

$$a^2 + 2 a b + b^2 = c^2 + 2 a b$$

$$a^2 + b^2 = c^2 \quad \text{q e d.}$$



Zweifellos erfordert die Lösung der Aufgabe Intelligenz (s.o. Def.). Bildet die Vergegenständlichung dieser uralten Methode also ein ›**KI-System**‹?

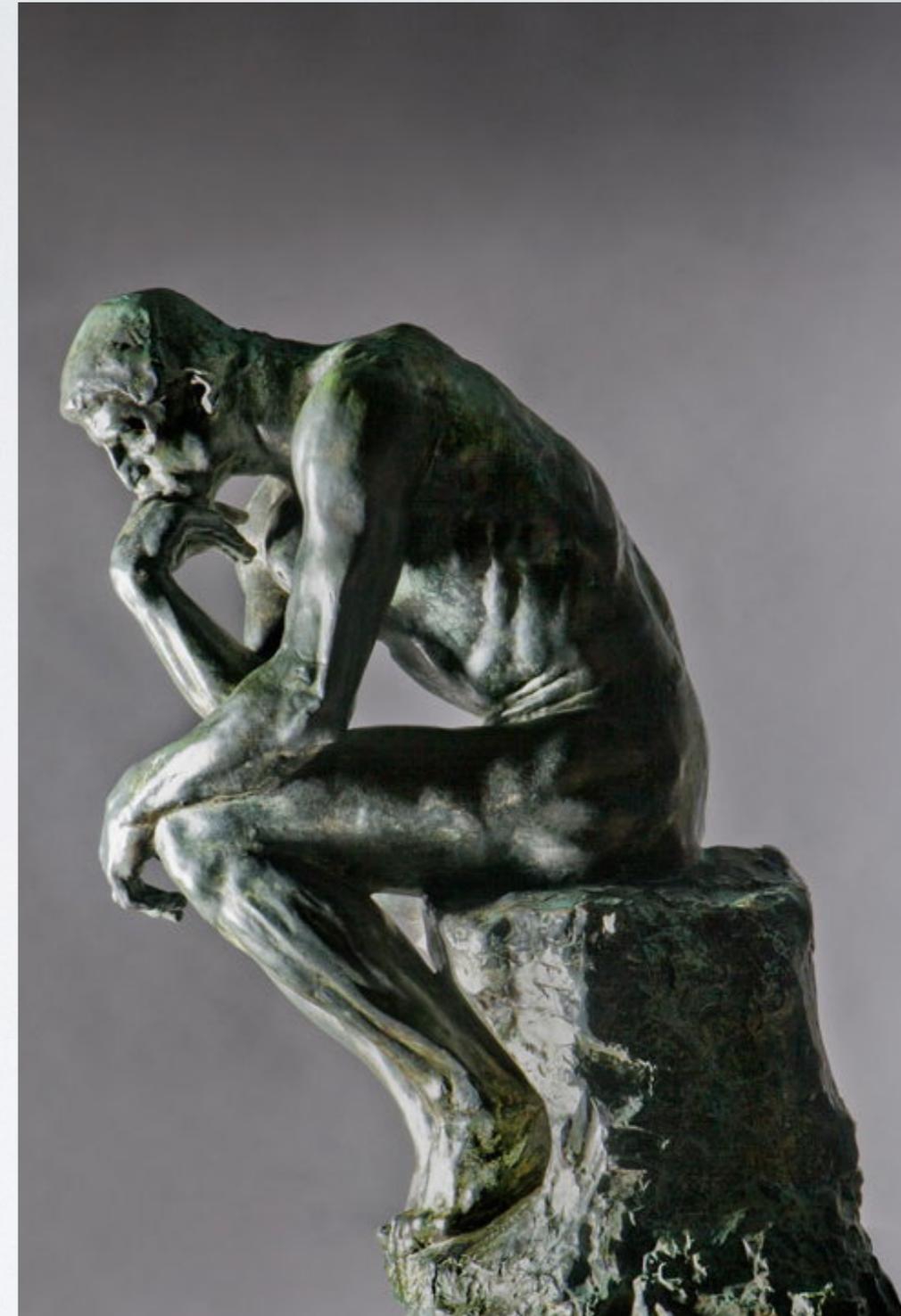
AlphaGo: Maßlos übertriebene Folgerungen

Als durch **Regeln** vollständig definierte **mathematische Objekte** sind Spiele geradezu **prädestiniert** für die **Formalisierung** ihrer Spielverläufe. Es ist daher **a priori** klar, dass es Algorithmen geben muss, die menschlichen Spielern **überlegen** sind. **Erklärungsbedürftig** ist folglich **nur**, warum sie **erst jetzt** gefunden wurden; Gründe dafür sind:

- es bedarf der Entwicklung notwendigen **mathematischen Könnens**, um leistungsfähige **Heuristiken** (z.B. MCTS) zur algorithmischen Bewältigung kombinatorischer Optimierung (oftmals NP-vollständige Probleme) zu finden,
- es muss **hinreichende Rechenleistung** – z.B. für das Training komplexer KNN – verfügbar sein.

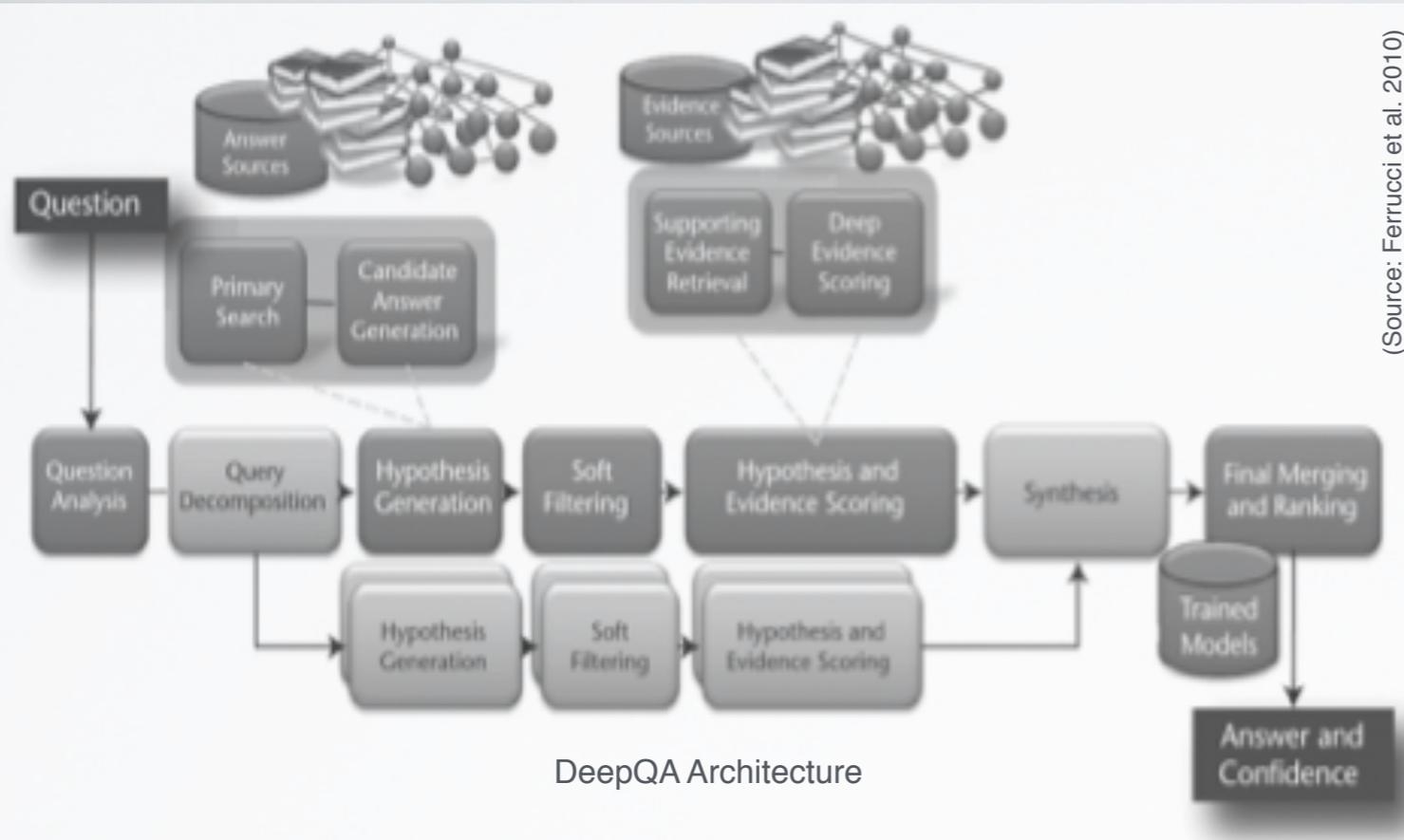
Infolge der **Trennung** von Planung und Ausführung gilt für **»KI«-Methoden** allgemein:

Auch **adaptive Systeme** sind algorithmisch gesteuert. Grob **irreführend** als »künstlich intelligent« bzw. »selbstlernend« bezeichnet, sind sie **stets & ausschließlich** das Ergebnis **methodischer Kompetenz** menschlicher Experten, deren **Können** und **natürliche Intelligenz** sie **vergegenständlichen** (letzteres gilt für Maschinen schon immer).



Rodin: Der Denker (Foto: Singer Museum)

IBM Watson: Gigantisches, aber zweifelhaftes Expertensystem



Software:

Datenbank mit ~100 GB enzyklopädischen Texten

UIMA Data Mining Framework für Verarbeitung unstrukturierter Daten

Apache Hadoop Framework für datenintensive Parallelverarbeitung

Hardware:

90 Power7 8-Kern-Prozessoren mit 3,5 GHz Taktfrequenz

Für den Sieg in **Jeopardy!** in 2011 wurden über **8 Jahre Entwicklungszeit** investiert (z.B. waren 2006 erst 15% der Antworten korrekt verglichen mit 95% bei Top-Teilnehmern).

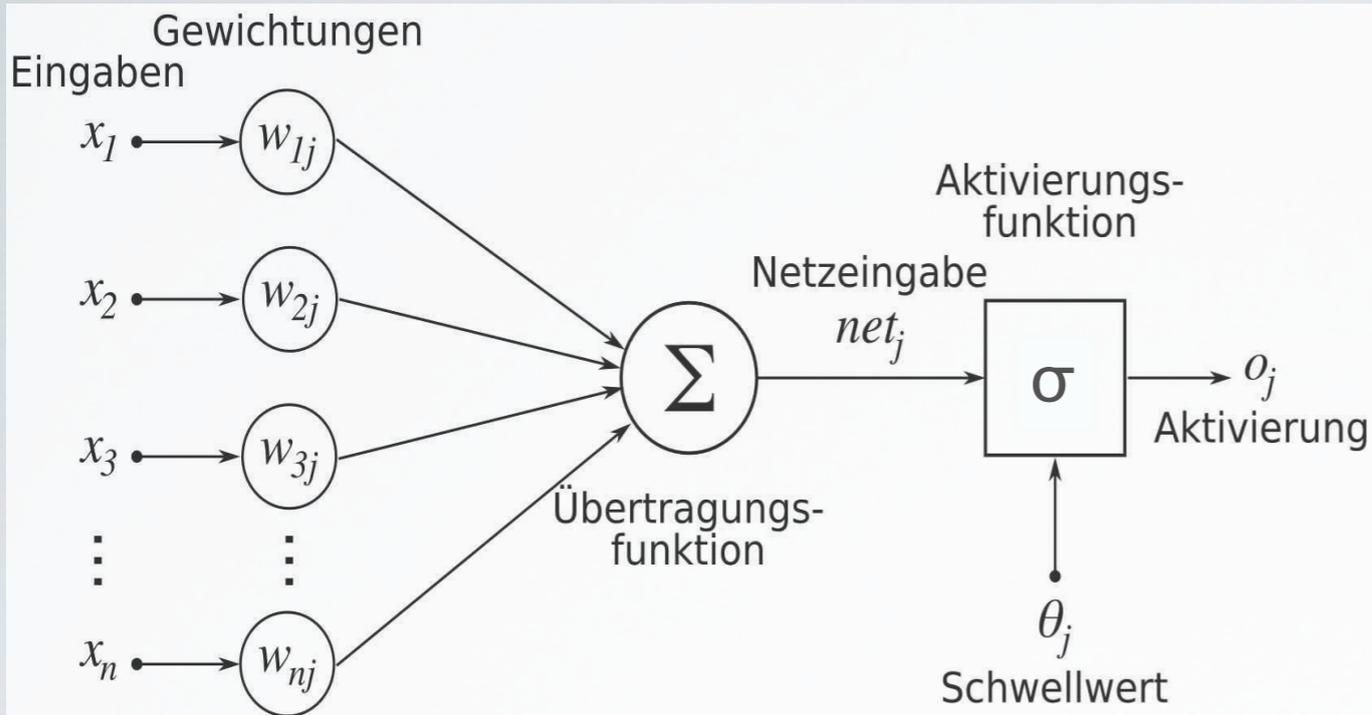
Die Verwendung der DeepQA Architektur in **anderen Wissensdomänen** – etwa in der medizinischen Krebsforschung und -therapie – erfordert, die **Wissensbasen** komplett **neu zu erstellen** und Verfahren der **Hypothesenbildung** anzupassen.

Nicht bewältigt werden dabei Probleme der **Validität der Ergebnisse**, insbesondere hinsichtlich der Beurteilung von Qualität & Repräsentativität der Daten.

So wurde Watson beim weltgrößten Krebsforschungszentrum Anderson in Houston wegen erwiesener **Unbrauchbarkeit gekündigt** (Lehrgeld: 62 Mio. USD).

(Quelle: Health News Review 2017)

»Deep Learning«: Künstliche Neuronale Netze (KNN)

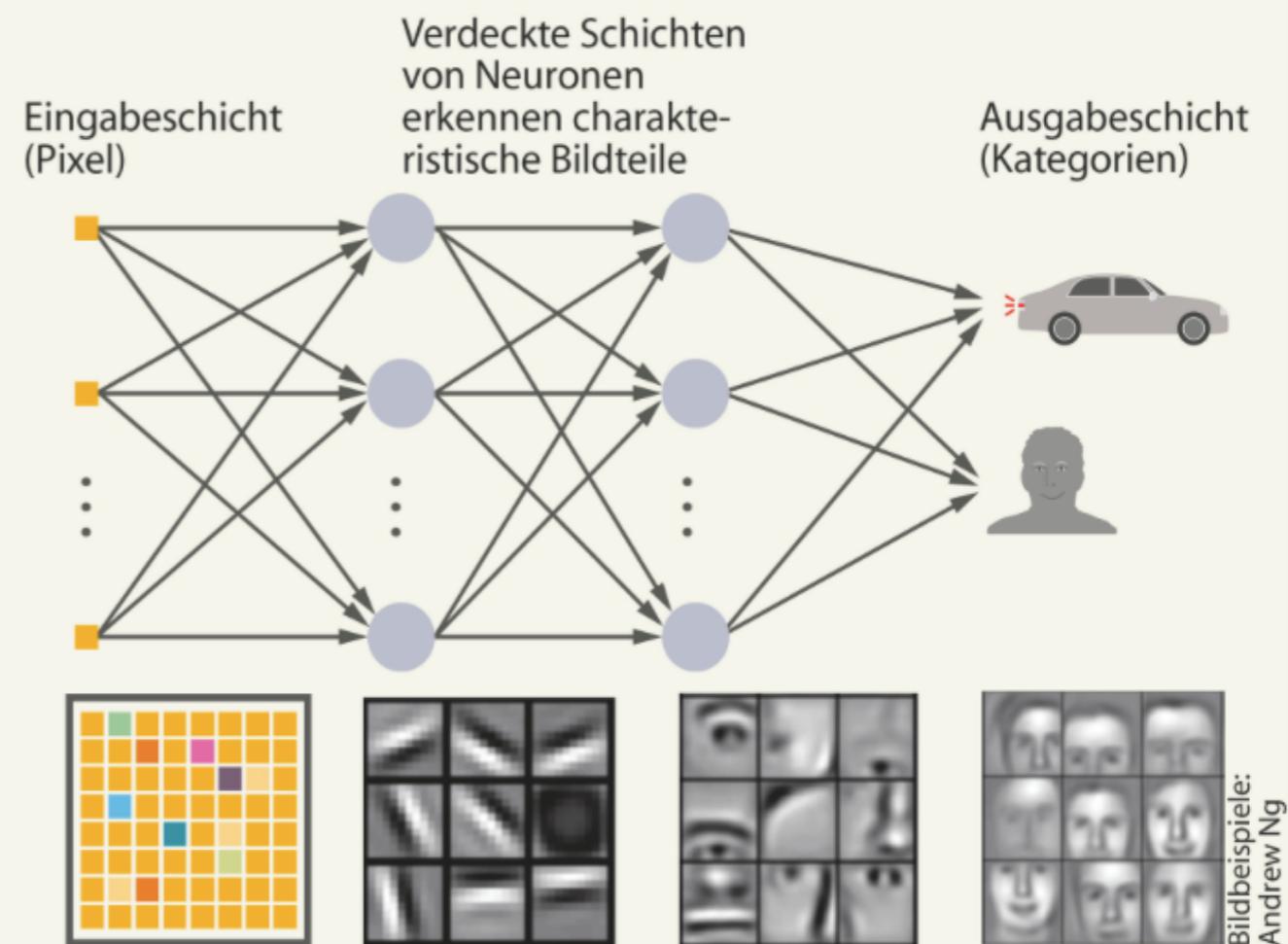


Berechnungsfunktionen eines **Netzknotens** j

KNN (speziell sog. Convolutional NN) erweisen sich seit langem als besonders geeignet für **Objektklassifikation** & Mustererkennung. Ein Netzwerk wird nicht programmiert, sondern nur **strukturiert** und mit ca. 10^6 Beispielen für die Aufgabe **trainiert** (algorithmisch gesteuerte Änderung der w_{ij}).

Die **Performanz** verdankt sich infolge **theorielosen Probierens** allein der **Erfahrung & Kreativität**, dem **Können** der Entwickler, sowie der exponentiell gesteigerten **Rechenleistung**.

Intelligent sind die Entwickler, **nicht** das KNN.



(Quelle: c't 6/2016)

»Deep Learning«-Probleme: schwindender Gradient & Schrittweite

Für die Bestimmung der **Gewichte** $W_j := W_j - \eta \nabla_w L$ eines **KNN** ist die Berechnung des **Gradienten** $\nabla_w L(N(x))$ einer Nutzen- oder **Verlustfunktion** $L(N(x))$ erforderlich (mit $N(x)$ als Netzwerkausgabe, $L(N)$ als z.B. über alle Trainingsbeispiele summierter euklidischer Distanz & der Lernrate η). Mit der Aktivierungsfunktion $\sigma(x)$ lässt sich – am einfachen Beispiel eines dreischichtigen KNN – der **Backpropagation**-Algorithmus und das Problem des **schwindenden Gradienten** aufzeigen:

$$N(x) = W_1 \cdot \sigma(\overbrace{W_2 \cdot \sigma(W_3 \cdot x)}^{N_2})$$

N_3

Die Bestimmung des Gradienten von L erfordert die Bildung der **Ableitung** von $N(x)$ als Verkettung mehrerer Funktionen nach der **Kettenregel**:

$$\begin{aligned}\frac{dL}{dW_1} &= \sigma(N_2) \cdot \frac{dL}{dN} \\ \frac{dL}{dW_2} &= \sigma(N_3) \frac{d\sigma(N_2)}{dN_2} \cdot W_1 \cdot \frac{dL}{dN} \\ \frac{dL}{dW_3} &= x \cdot \frac{d\sigma(N_3)}{dN_3} \cdot W_2 \cdot \frac{d\sigma(N_2)}{dN_2} \cdot W_1 \cdot \frac{dL}{dN}\end{aligned}$$

Dabei häufig wiederkehrende Faktoren sind während des Trainings oft < 1 , sodass insgesamt das Ergebnis gegen Null tendiert – daher der »schwindende Gradient« und die **Schwierigkeit**, vielschichtige KNN zu **trainieren**.

Zudem ist es **schwierig**, eine variable, **situativ passende Schrittweitensteuerung** zu realisieren.

Zur Erinnerung: **Prinzipielle Grenzen** formaler Systeme

Selbst die äußerst formalisierte **Mathematik** widersetzt sich ihrer vollständigen Algorithmisierung. Es ist erwiesenermaßen **unmöglich**,

- einen Algorithmus anzugeben, der **alle Sätze** eines formalen Systems **abzuleiten** und deren **Widerspruchsfreiheit** zu zeigen imstande ist (Gödel 1931);
- einen Algorithmus anzugeben, der von **jeder Formel** eines formalen Systems **entscheiden** kann, ob diese Formel ein wahrer Satz des Systems ist (Turing 1937).

Bezeichnenderweise beruht der Beweis von Gödel im Kern darauf, dass er als Mathematiker eine Formel im System so konstruiert, dass sie über einen durch ihn als wahr erkannten Satz aussagt, nicht beweisbar (ableitbar) zu sein.

Zur **mathematischen Fähigkeit** von Menschen gehört eben auch, dass sie über alles, was sie zu formalisieren vermag, durch Nachdenken über die Formalisierung mittels **abduktiven Schließens** über sie **hinaus gelangen** kann.

Die Abfolge von Operationen eines Algorithmus gibt zwar Auskunft auf die Frage, **was** abläuft; sie beantwortet aber nicht die Frage nach deren Sinn oder Bedeutung, **warum** sie so abläuft. Operationsfolgen sagen **nichts über sich selbst** aus, etwa ob sie korrekt oder gebrauchstauglich sind. So ist auch die Frage, ob ein Algorithmus **terminiert**, formal **nicht entscheidbar**.

Irrige Ontologie von **Mensch** und **Maschine**

Infolge von Begriffs-**Missbrauch** ist die Vision »**autonomer Systeme**« auf Sand gebaut.
»Die Philosophie ist ein Kampf gegen die Verhexung unseres Verstandes durch die Mittel unserer Sprache.«

(Wittgenstein 1984 PU: 109)

Mensch (lebendig)

Sich per **Autopoiese** durch Stoffwechsel und Kommunikation **selber machend**.

Autonom (per selbstbestimmter Regeln).

Handelt **intentional** (kontingent).

Ist sprachbegabt, **reflektiv lernfähig**.

Lebendiges **Arbeitsvermögen**:

Können (implizites Wissen, Erfahrung, situierte Urteilskraft & Handlungskompetenz), **verausgibt** & **reproduziert** sich im Handeln.

(Semiotische) Maschine (tot)

Wissensbasiert **für** bestimmte **Zwecke gemacht**.

Automatisch (auto-operational, selbsttätig).

Verhält sich **kausal** determiniert;

ggf. algorithmisch gesteuert **Umwelt-adaptiv**.

Algorithmisch determiniertes **Verhalten**:

Setzt **Formalisierung** von Zeichenprozessen voraus, muss für Praxis **angeeignet** & **organisatorisch eingebettet** werden.

Die **irreführende Metaphorik** über semiotische Maschinen, »**als ob**« diese **wie** Menschen wären – »autonom«, »selbstorganisiert«, »intelligent«, »smart«, »selbstlernend«, »selbstheilend« etc. –, **ignoriert** in grob **reduktionistischer** Weise die fundamentalen **Unterschiede**.

Dadurch wird kompetentes **Handeln** von **Menschen** auf **algorithmisch** gesteuertes **Verhalten** von **Maschinen reduziert**; zugleich entstehen **Illusionen** über deren tatsächliche **Leistungsfähigkeit**.

Im KI-Wahn vom Eigenleben der Maschinen äußert sich die

»Macht der Machwerke über die Machenden« (W.F. Haug) bzw. »prometheische Scham« (G. Anders).

Auf die **Entwicklungsperspektive** kommt es an

Die **Gestaltung soziotechnischer Systeme** muss sich an Erfordernissen menschlichen Handelns und der Entfaltung von Arbeitsvermögen orientieren, um gute Arbeit, Produktivität & Innovation zu ermöglichen.

Technikzentrierte Perspektive:

AI (Artificial Intelligence)

»Smart machines«, »autonome Agenten«

(»Intentional stance«; Dennett 1987; Shoham 1993, Wooldridge 2002)

MAS ersetzen Menschen in der Produktion, Flexibilität angestrebt, aber nur Adaptivität statt »Lernfähigkeit«.
Arbeitsvermögen wird nachgeahmt und partiell ersetzt.

Verbleibende Restarbeit:

»Ironies of automation« (Bainbridge 1983, Baxter 2012),

Verlust praktischer Handlungskompetenz, hilflose »Bediener«;
intransparente & nicht erwartungskonforme Systeme.

Beispiele »**normaler Katastrophen**« (Perrow 1989):

Ariane V: Totalverlust beim Erstflug 1996;

HFT-Debakel: Flash-Crash 2010, Knight»mare« 2012;

»Stalk«-Unfälle: Air France AF 447 Südatlantik 2009,

Continental Connection flight Buffalo 2009 (Carr 2013).

Tesla »Autopilot«-Versagen 2016-18

Praxistheoretische Perspektive:

IA (Intelligence Amplification)

»Things (machines) that make us smart«

(Norman 1993; vgl., Engelbart 1962, Ehn 1988, Winograd 1996)

IT unterstützt lebendige Arbeit, ermöglicht Entfaltung von Arbeitsvermögen mit gesteigerter Produktivkraft und Innovationsfähigkeit durch:

Menschengerechte, reflexive Gestaltung von Arbeitsaufgaben, technischen Arbeitsmitteln und Interaktionsformen; Fokus auf Entfaltung praktischer Handlungskompetenz.

Methodisch saubere Analyse von **Daten aus CPS** lässt sich z.B. nutzen als Grundlage für **interaktive Assistenzsysteme**, als selbsterklärende, **gebrauchstauglich gestaltete Hilfsmittel** zur Optimierung von Prozessen, zu datengestützter Diagnose von Anlagen oder zu effektiver Simulation bei der Steuerung von Prozessen.

Trotz **AI**-Versuchen: Erfolgreiche Computernutzung beruht hauptsächlich auf **IA**-Gestaltungsperspektive (PC- & Internet-basierte SW-Werkzeuge für Bearbeitung & Austausch digitaler Gegenstände etc.).

Literatur

- Autorengruppe (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, Oxford (AR): Future of Humanity Institute u.a. 02/2018, <https://arxiv.org/pdf/1802.07228.pdf>
- Bainbridge, L. (1983): Ironies of Automation, in: Automatica 19 (6), 775-779
- Baxter, G.; Rooksby, J.; Wang, Y. & Khajeh-Hosseini, A. (2012): The Ironies of Automation ... still going strong at 30?, in: Turner, P. & Turner, S. (ed.): European Conference on Cognitive Ergonomics, ECCE '12, Edinburgh, United Kingdom, 65-71
- Browne, C. et al. (2012): A Survey of Monte Carlo Tree Search Methods, in: IEEE Transactions on Computational Intelligence and AI in Games 4 (1), 1-49
- Carr, N. (2013): All Can Be Lost: The Risk of Putting Our Knowledge in the Hands of Machines, The Atlantic No. 11
- Dennett, D.C. (1987): The Intentional Stance. Cambridge (MA): MIT Press
- Ehn, P., 1988: Work-Oriented Design of Computer Artifacts, Stockholm: Arbetslivscentrum
- Ferrucci, D. et al. (2010): Building Watson: An Overview of the DeepQA Project, AI Magazine Fall 2010, 59-79
- Hofstadter, D.R., 1979: Gödel, Escher, Bach. An Eternel Golden Braid, New York: Vintage Books
- Norman, D. A., 1993: Things that Make Us Smart. Defending Human Attributes in the Age of the Machine, Reading, (MA): Addison-Wesley
- Perrow, C., 1989: Normale Katastrophen. Die unvermeidbaren Risiken der Großtechnik, Frankfurt/M: Campus
- Russel, S. & Norvig, P. (2009): Artificial Intelligence: A Modern Approach, 3rd. ed., Essex (GB): Pearson
- Silver, D. et al. (2016): Mastering the Game of Go with Deep Neural Networks and Tree Search, in: Nature 529, S. 484-489
- Shoham, Y. (1993): Agent-oriented Programming, Artificial Intelligence 60, 51-92
- Wick, C. (2017): Deep Learning, Informatik Spektrum 40 (1), 103-107
- Winograd, T. (Ed.), 1996: Bringing Design to Software, Reading (MA): Addison-Wesley
- Wooldridge, M.J., 2002: An Introduction to Multiagent Systems, New York: Wiley
- Yuandong, T. & Yan, Z. (2016): Better Computer Go Player with Neural Network and Long-term Prediction, arXiv:1511.06410v3, <https://arxiv.org/pdf/1511.06410.pdf>