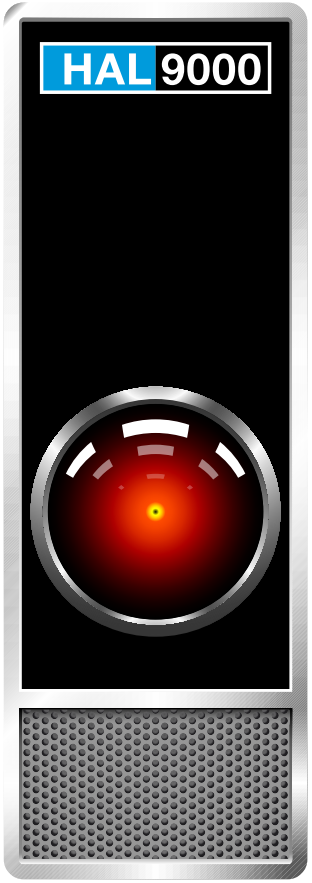
KI – Kybernetik – Maschinenethik

Versuch einer Reflexion jenseits von Hype und Horror

Beitrag zum Darmstädter Ontologenkreis
12. Januar 2022

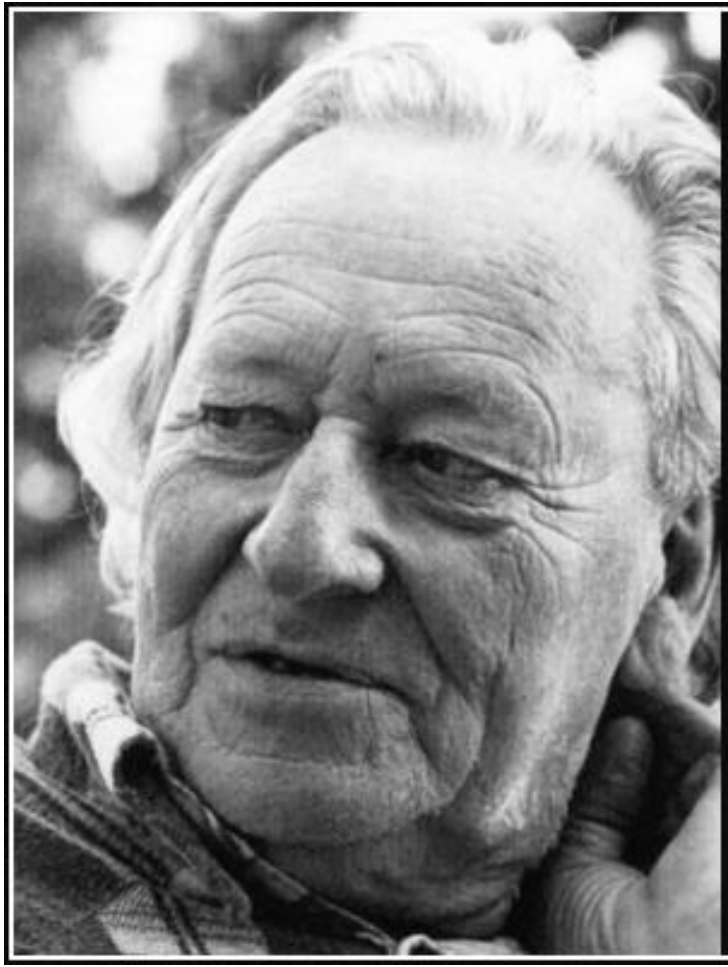
Dr. Joachim Paul

Zuerst … ein Geburtstag ...



HAL 9000, der Bordcomputer, die KI des Raumschiffs Discovery aus Stanley Kubricks Film *"2001: Odyssee im Weltraum"* wurde heute – laut Filmhandlung – 30 Jahre alt.
Der Film selbst ist 50 Jahre alt.

Arthur C. Clarke:
HAL = **H**euristically programmed **AL**gorithmic computer



The major problems in the world are the result of the difference between how nature works and the way people think.

— Gregory Bateson —

AZ QUOTES

KI – Kybernetik – Maschinenethik

Der Begriff Künstliche Intelligenz und seine Anverwandten tragen schon seit Jahren alle Charakteristika eines gehypten Narrativs. Dabei wurden und werden einige Vorannahmen und Hypothesen als Tatsachen behandelt und entweder nur unzureichend reflektiert oder gar völlig vernachlässigt.

Der Beitrag versucht, über die Wiedereinbeziehung von Erkenntnissen insbesondere aus der Kybernetik Standpunkte jenseits der Polarität von Hype und Horror zu entwickeln, die den menschlichen Konstrukteuren und Nutzern von KI ihre zwischenzeitlich abhanden gekommene Souveränität zurückgeben könnten.

(So man das will …)

Dr. Joachim Paul, Jan. 2022

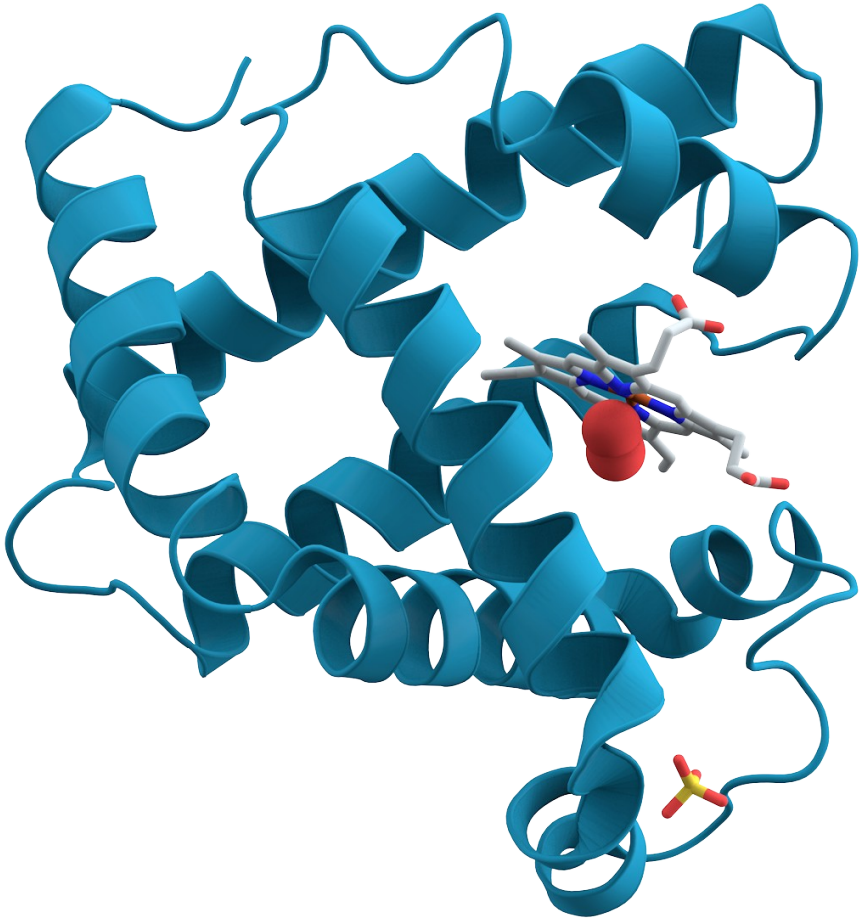
KI – Kybernetik – Maschinenethik

Leistungen – Auszüge und Schlaglichter

KI – Kybernetik – Maschinenethik

Leistungen – Auszüge und Schlaglichter

1. AlphaFold (Deepmind (Google))
2. Exominer (NASA, Pleiades, Deep Learning, 301)
3. GPT-3 (OpenAI (Musk, Microsoft, etc.))

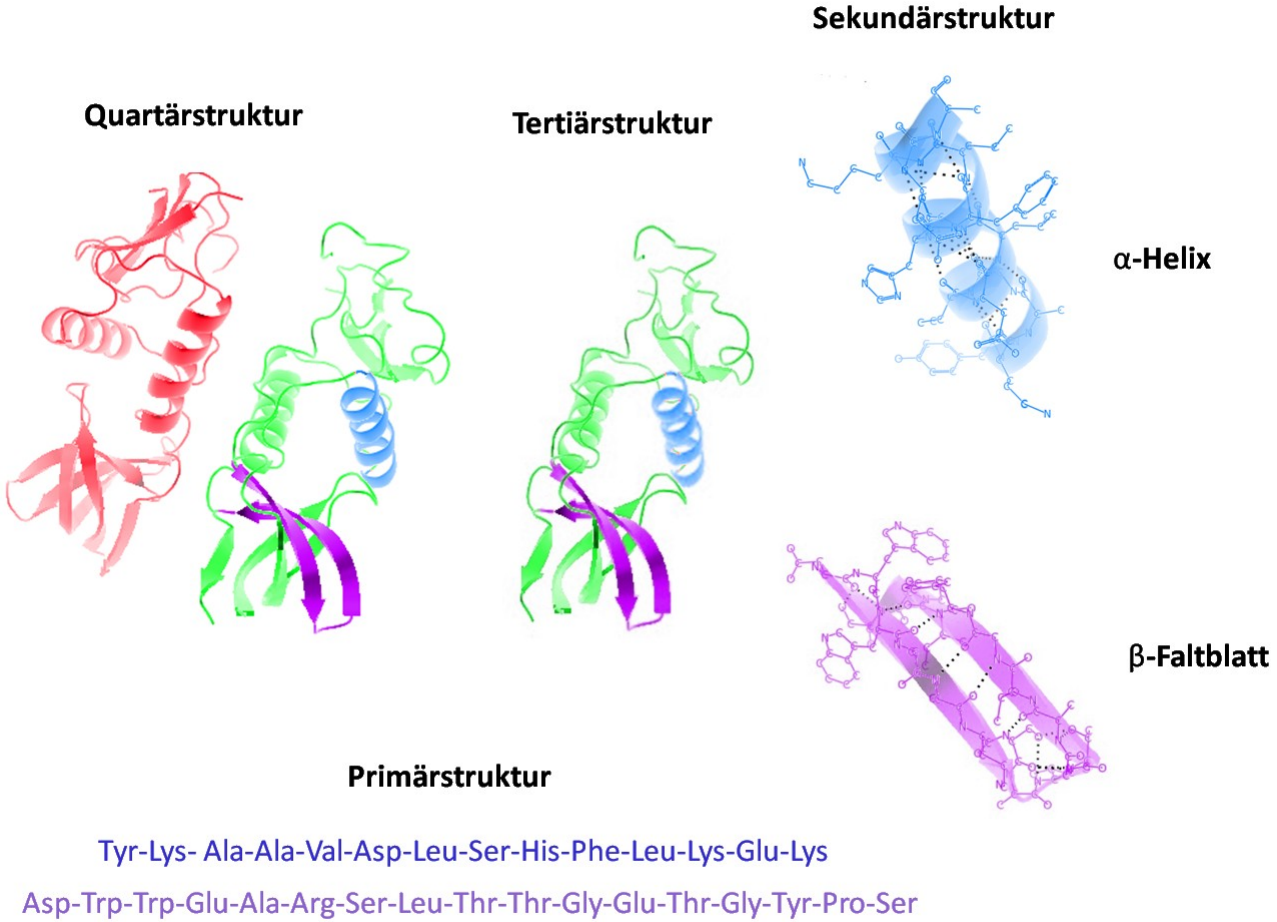


Myoglobin im Bändermodell (Sekundärstruktur), bzw. Stäbchenmodell (für die aktive Porphyringruppe, auch Häm genannt)

Sauerstofftransporter von Zellmembran zu Mitochondrien vorw. in Muskelzellen, mglw. auch Speicher

Einkettiges Protein aus 153 Aminosäuren, 8 α-Helices, Masse: 17053 Dalton (1 Da = 1/12 der Masse von C^{12})

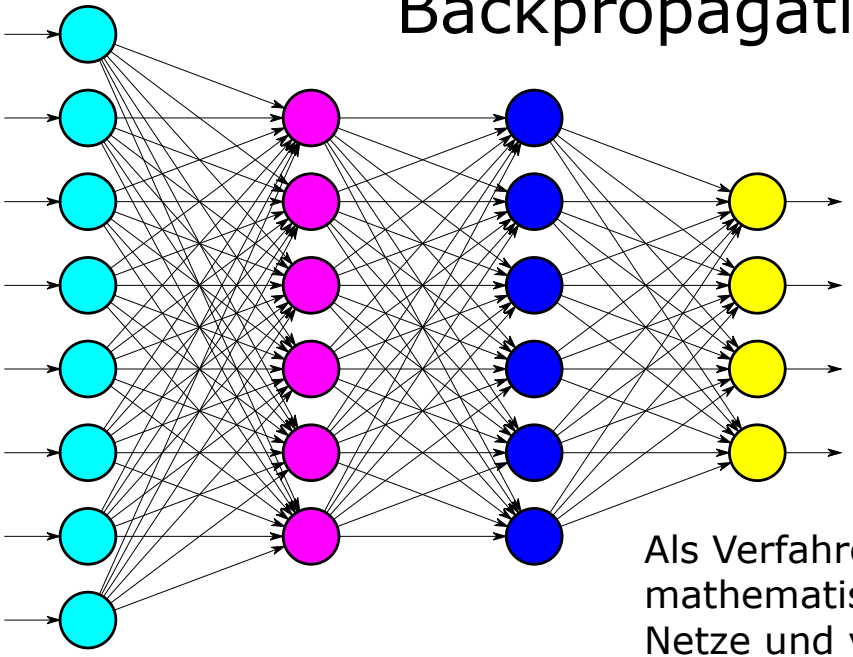
Erste Proteinstrukturaufklärng eines Proteins überhaupt mit „Röntgenstrukturanalyse" durch John Kendrew 1958, (Myoglobin vom Pottwal) Nobelpreis für Chemie 1962 zus. mit Max Perutz (Hämoglobin 1959)

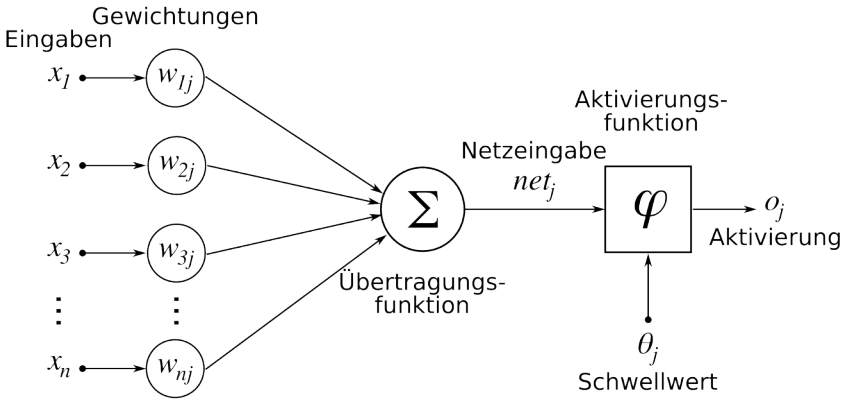


Alphafold

Input: Primärstruktur
Output: Sekundär- bis Quartärstruktur

Backpropagation / Deep Learning





Als Verfahren des maschinellen Lernens ist Backpropagation ein mathematisch fundierter Lernmechanismus künstlicher neuronaler Netze und versucht nicht, tatsächliche neuronale Lernmechanismen biologistisch zu modellieren.

Problem: Einstellung der Gewichtungen bei mehr als einem hidden layer.
Ausweg: LSTM, long short-term memory neurons, Schmidhuber.

KI – Kybernetik – Maschinenethik

Problemkontexte

Medien, Gesellschaft, Individuum, Arbeitswelt

Dr. Joachim Paul, Jan. 2022

Vorab: Was ist eine Maschine?

Definition, herkömmlich:

1. Sammelbezeichnung für zweckorientierte technische Vorrichtungen verschiedenster Art und Größe mit i.d.R. beweglichen Teilen.

2. Teil des Sachanalagevermögens.

Aus: Gabler Wirtschaftslexikon, online 2022:
https://wirtschaftslexikon.gabler.de/definition/maschine-40202

Maschine, eine möglichst allgemeine Definition

Eine Maschine (altgr. μηχανή, mēchanḗ, dt. etwa Werkzeug, künstliche Vorrichtung, Mittel) ist in einem sehr abstrakten und prinzipiellen Sinn eine technisch verfertigte Vorrichtung, ein Instrument zur Herstellung und/oder Aufrechterhaltung von materiellen oder immateriellen Relationen.
Physikalisch gesehen benötigt eine Maschine immer Energie. Man unterscheidet Maschinen, die Energie und/oder Materie prozessieren und Maschinen, die Signale oder Informationen prozessieren.

Joachim Paul, *Eine Krise unserer Rationalität? - Zur Dialektik der Beziehungen zwischen Ökonomie, Umwelt und Technik*, in: Heinz-J. Bontrup, Jürgen Daub (Hg.), *Digitalisierung und Technik – Fortschritt oder Fluch? Perspektiven der Produktivkraftentwicklung im modernen Kapitalismus*, Köln 2021, S. 74-113

Vorschlag inspiriert d.: Lewis Mumford, *Mythos der Maschine*; Martin Burckhardt, *Theorie der Maschine*

„Werbeblock“ ☺

Heinz-J. Bontrup / Jürgen Daub (Hg.)

Digitalisierung und Technik – Fortschritt oder Fluch?

Perspektiven der Produktivkraftentwicklung im modernen Kapitalismus

Paperback, 321 Seiten

Erschienen (November 2020)

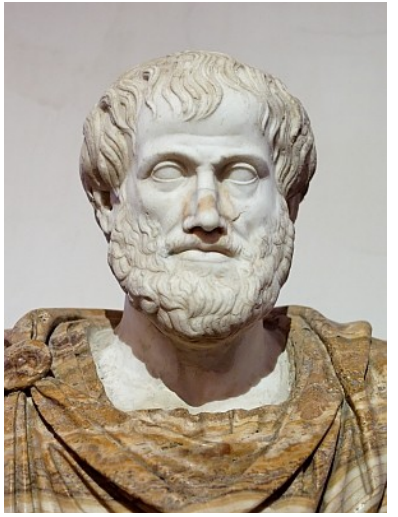
Mit Beiträgen von Gustav Bergmann, Peter Brödner, Florian Butollo, Alex Demirović, Rainer Fischbach, Joachim Paul und Rudi Schmiede.

Die erste Maschine I

Die erste Maschine II

Die Megamaschine „Gesellschaft“ produziert Artefakte.

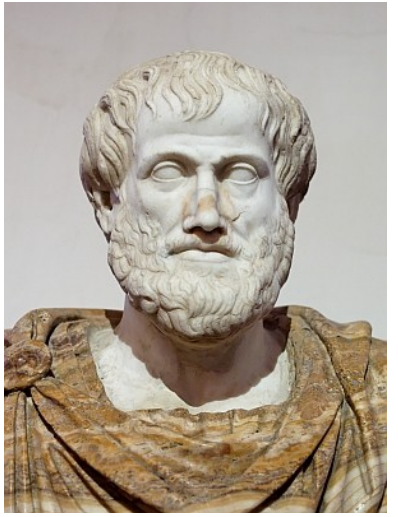
Werner Hamacher zur Technik bei Aristoteles



"Dass die Natur mit der Technik über sich selber hinauskommt, besagt aber wiederum Zweierlei, nämlich zum Ersten, dass sie dahin kommt, wo sie zuvor nie gewesen ist, und zum Zweiten, dass sie erst dort, wo sie nie war, bei sich als Natur ankommen kann. Technê ist somit diejenige Verfertigung der Natur, in der die Natur ganz zu sich als einem von ihr Unterschiedenen zurückkommt. Dieses von der Natur sowohl Unterschiedene, dass zugleich sie selbst in ihrem Wesen ist, heißt für Aristoteles ihr Telos."

Werner Hamacher, *Technik, Löffelheit, gedachter Verstand*
– Vortrag Bochumer Kolloquium Medienwissenschaften 25.05.2011
Video-URL:
https://www.ruhr-uni-bochum.de/bkm/archivseiten/23_hamacher.html
Timecode: 00:12:43

Aristoteles, Übersetzungen des Originals



"Das Verhältnis zwischen Natur und menschlichem Bemühen ist an der bekannten Stelle in Phys. II 8, 199a15 ff. in die Formel gekleidet: „Die 'Technik' bringt im allgemeinen das zur Vollendung, was die Natur nicht imstande ist, zu Ende zu führen, im übrigen ahmt sie die Natur nach."
Es geht an dieser Stelle zwar mehr um die Produkte von Natur und technê, aber "Entsprechendes gilt auch für den Techniten und Philosophen. Er ist der verlängerte Arm der Physis."

Otfried Höffe, *Aristoteles: Die Nikomachische Ethik,* 4. Aufl., Berlin/Boston 2019

"Phys. II 8 Und könnte umgekehrt das Natürliche nicht nur durch Natur, sondern auch durch Kunst entstehen, so würde es eben so werden, wie es von Natur ist. Wegen des Einen also kann das Andere sein. Und überhaupt vollendet die Kunst theils was die Natur nicht zu vollbringen vermag, theils ahmt sie sie nach."

Aristoteles, Gesammelte Werke, Translator: Alfred Gudeman, Adolf Lasson, J. H. von Kirchmann, C. H. Weiße, e-artnow 2016

KI – Kybernetik – Maschinenethik

Statements und Thesen

– eine sowohl höchst heterogene
als auch sträflich unzureichende Sammlung

Dr. Joachim Paul, Jan. 2022

KI – alt und langsam

Charlie Stross:
Auch Unternehmen sind künstliche Intelligenzen, i.d.R. Maximierer von Irgendetwas

"Tesla is a battery maximizer."

Aus: *Dude, you broke the future!*, By Charlie Stross
keynote speech at the 34th Chaos Communication Congress, 34C3 in Leipzig, Dec 2017
http://www.antipope.org/charlie/blog-static/2018/01/dude-you-broke-the-future.html

André Gorz über Systemzusammenhänge

Die Wissenschaft hat dem Kapital […] *"den Weg bereitet, indem sie die anschauliche Welt ausgeklammert und die Wirklichkeit als ein der reinen Logik des Kalküls gehorchendes, nur mathematisch denkbares System von Relationen erfasst hat. Die entsinnlichten, elektronisch schaltbaren mathematischen Denkprozesse haben der politischen Ökonomie die Mittel beschafft, mit rechnerischen Realabstraktionen die gesellschaftlichen Verhältnisse zu gestalten. Sie haben zu einer vom lebendigen Erfahrungswissen abgespaltenen, den Sinnen unzugänglichen Systemwelt geführt. In ihr erscheint der Mensch als ein überfordertes, antiquiertes, heimatloses Wesen."*

André Gorz, *Wissen, Wert und Kapital - Zur Kritik der Wissensökonomie*, Zürich 2004

Anmerkung: André Gorz über Systemzusammenhänge

Das Zitat weckt Assoziationen in Richtung des (u.a.) Husserl- u. Heidegger-Schülers Günther Anders und seiner Antiquiertheit des Menschen. Tatsächlich nimmt Gorz an anderen Stellen Bezug auf Anders und zitiert aus dessen Werk *Die Antiquiertheit des Menschen*. Band II: *Über die Zerstörung des Lebens im Zeitalter der dritten industriellen Revolution.,* München 1980

Anders: *"Der Mensch wird nebengeschichtlich.*"

spontan assoz. Frage: "*Ja wer oder was ist dann hauptgeschichtlich?*"

Bezugnahmen u.a. auf Rainer Fischbach, Erich Hörl, Peter Sloterdijk, etc.

André Gorz über KI

"Er (der Mensch) braucht chemische und elektronische Prothesen, um der technischen Umwelt gewachsen zu sein. Das Projekt »Künstliche Intelligenz/Künstliches Leben« soll die biologische Begrenztheit des Menschen beseitigen. Die [...] reichlich zitierten Pioniere der künstlichen Intelligenz – Minsky, Moravec, Kurzweil, de Garis usw. – machen keinen Hehl aus ihrer Verachtung für die menschliche »Fleischmaschine«. Die Natur, denken sie, habe der Gattung Mensch die Fähigkeit gegeben, sich selbst zugunsten post-biologischer Lebens- und Intelligenzformen aufzuheben, ja sich in digitaler Form als unsterblicher Geist im Weltall aufzulösen."

André Gorz, *Wissen, Wert und Kapital - Zur Kritik der Wissensökonomie*, Zürich 2004

Jürgen Schmidhuber über KI

"Die Grundlagen der heute populärsten KI-Algorithmen stammen alle aus dem letzten Jahrtausend. Schon seit über drei Jahrzehnten gibt es lernende Maschinen mit so etwas wie Emotion und Selbstbewusstsein: Diese stecken sich ihre eigenen Ziele, statt nur sklavisch dem Menschen zu dienen.
Denkende KIs werden Menschen spätestens in ein paar Jahrzehnten bei fast allem übertreffen, das ihnen wichtig ist."
[…]
"Seit Jahrmilliarden scheint der Trend offensichtlich: Das Universum wird komplexer. Diese Entwicklung scheint unaufhaltsam fortzuschreiten. Nun macht das Universum seinen nächsten Schritt hin zu höherer Komplexität."

Christine Schulthess, Interview mit Jürgen Schmidhuber, *Die Menschheit ist bloss Steigbügelhalter für etwas Grösseres*, SRF 05.01.2022

Anmerkungen zu Schmidhuber über KI

Die abnehmende biologische Diversität auf der Erde spricht jedoch eher für abnehmende Komplexität.

Von einer Komplexitätszunahme kann bestenfalls gesprochen werden, wenn wir unsere Technik und unser Zusammenspiel mit ihr in die Betrachtung mit einbeziehen.

Roberto Simanowski: Todesalgorithmus

"Die Selbstentmachtung des Menschen durch sein eigenes Geschöpf wäre die Vollendung eines Aufbruchs, der mit dem Griff nach der Erkenntnis begann. Überlässt der Mensch die Regelung seiner Angelegenheiten der KI, wie es ansatzweise schon geschieht, wenn Apps ihn von A nach B navigieren und Algorithmen die Partnerwahl übernehmen, kehrt er im Grunde an den Anfang zurück: als er noch unfähig war, Gut und Böse zu unterscheiden.
Diese Rückkehr enthebt ihn nicht nur der Notwendigkeit, selbst zu entscheiden, sondern auch der Gefahr, sich dabei zu irren. Ist es die Heimkehr ins Paradies? Wäre ein solches Paradies wünschenswert? Die Antwort hängt davon ab, worin man den Sinn des Lebens sieht und wieviel Autonomie, Nudging und Social-Scoring darin enthalten sein soll."

Roberto Simanowski, *Todesalgorithmus*, Wien 2020, S.11f

Roberto Simanowski: Todesalgorithmus

"Immerhin verfügt der Bordcomputer über die Fahrerfahrung aller Computer, verarbeitet viel mehr Information viel schneller als der Mensch, wird niemals müde, fährt nie betrunken und textet nicht am Lenkrad. Es wäre moralisch unverantwortlich, selbstfahrende Autos nicht einzuführen. Zugleich gilt als sicher, dass es weiterhin Unfallsituationen geben wird, bei denen Todesopfer nicht vermeidbar, sondern nur wählbar sind. Das führt zum ethischen Problem, die Algorithmen der selbstfahrenden Autos mit einer spezifischen Entscheidungsmoral ausstatten zu müssen, die unvermeidlich dem Grundprinzip der deutschen Verfassung widerspricht. Die Unantastbarkeit der Menschenwürde ist in Gefahr, wenn die Technik verlangt, Abwägungen, die sich philosophisch verbieten, für den Ernstfall zu programmieren."

- technisches Triagieren … - Stichwort Kobayashi Maru Test

Roberto Simanowski, *Todesalgorithmus*, Wien 2020, S.16

Yuval Noah Harari – Homo Deus

"Organismen sind Algorithmen" […]
"Wie Naturwissenschaftler in den letzten Jahrzehnten gezeigt haben, sind Emotionen keineswegs irgendein rätselhaftes seelisches Phänomen, das allein dazu dient, Gedichte zu verfassen und Symphonien zu komponieren. Emotionen sind vielmehr biochemische Algorithmen, die für das Überleben und die Reproduktion sämtlicher Säugetiere von entscheidender Bedeutung sind.[…] Die Algorithmen, die Getränkeautomaten steuern, funktionieren über mechanische Getriebe und Stromkreise. Die Algorithmen, die Menschen steuern, funktionieren über Sinneswahrnehmungen, Emotionen und Gedanken. Und genau die gleiche Art von Algorithmen steuert Schweine, Paviane, Otter und Hühner."

Yuval Noah Harari, *Homo Deus – Die nächste Stufe der Evolution*, München 2017

Yuval Noah Harari – Homo Deus

"1. Sind Organismen wirklich nur Algorithmen, und ist Leben wirklich nur Datenverarbeitung?
2. Was ist wertvoller – Intelligenz oder Bewusstsein?
3. Was wird aus unserer Gesellschaft, unserer Politik und unserem Alltagsleben, wenn nichtbewusste, aber hochintelligente Algorithmen uns besser kennen als wir uns selbst?"

Yuval Noah Harari, *Homo Deus – Die nächste Stufe der Evolution*, München 2017

Yuval Noah Harari – Homo Deus

"Ein Kräuseln im Datenfluss"

"Wie wir in Kapitel 3 gesehen haben, ist durchaus zweifelhaft, ob sich das Leben wirklich auf Datenströme reduzieren lässt. Insbesondere haben wir im Moment keinerlei Vorstellung, wie oder warum Datenströme Bewusstsein und subjektive Erfahrungen erzeugen könnten. Es kann sein, dass wir in zwanzig Jahren über eine plausible Erklärung verfügen. Vielleicht finden wir aber auch heraus, dass Organismen gar keine Algorithmen sind."
[…]
"Selbst wenn der Dataismus unrecht hat und Organismen nicht nur Algorithmen sind, wird das den Dataismus nicht zwangsläufig davon abhalten, die Welt zu übernehmen."

Yuval Noah Harari, *Homo Deus – Die nächste Stufe der Evolution*, München 2017

Richard David Precht

"Kein Zweifel: Das Virus weckt die Welt aus ihrem techno-topischen Schlummer. Die wirkliche Wirklichkeit ist nicht digital. Doch in den Visionen des Silicon Valley (Bill Gates einmal ausgenommen) gibt es keine unberechenbare Natur, nur eine permanent fortschreitende Technisierung von allem.
Jede Kurve geht exponentiell nach oben: Schneller, Höher, Weiter und Mehr! Beschleunigung, der Fetisch der Gelangweilten, ist alternativlos, bedingungslose Expansion, der Fetisch der Wertfreien, ebenso. Für die Geistesgegenwart aber sollten künftig mehr und mehr kluge Maschinen sorgen. Wie sehr haben sie sich geirrt!"

Richard David Precht, *Künstliche Intelligenz und der Sinn des Lebens*, München 2020, S.7

"Doch gerade dieses »In-der-Welt-Sein« ist, wie Martin Heidegger in der ersten Hälfte des 20. Jahrhunderts gezeigt hat, elementar für alles menschliche Erleben. Von hier aus unterscheiden Menschen (wie andere Lebewesen auch), was für sie relevant ist und was nicht. Eine riesige Menge stummen Wissens durchzieht unseren Alltag, bestimmt unsere Handlungen und unsere Sprache. Bedeutungen werden nicht logisch erschlossen, sondern dem Kontext abgelauscht. Unser Denken hat einen feinen, gesamtkörperlichen Sinn für Stimmungen, Zwischentöne und komplexe Zusammenhänge. Jedes Thema erscheint uns in einem Horizont von persönlichem und kulturellem Vorwissen. Der US-amerikanische Philosoph Hubert Dreyfus [...] wurde seit den Sechzigerjahren nicht müde, diesen Wissensschatz der Philosophie seinen oft unverständigen Kollegen aus der Informatik nahezubringen."

Richard David Precht, *Künstliche Intelligenz und der Sinn des Lebens*, München 2020, S.29f

Zu Hubert L. Dreyfus

"Im Rückgriff auf die Hermeneutik Heideggers kritisierte Hubert Dreyfus in „Cybernetics as the Last State of Metaphysics" bereits 1968 Marvin Minskys Statement „There is no reason to suppose machines have any limitations not shared by men", indem er auf einen infiniten Regress bzw. einen circulus vitiosus hinwies (vgl. Dreyfus 1968; Kaehr 1980). Dreyfus' Argumentationsgang kann wie folgt nachgezeichnet werden: Nehmen wir an, dass unsere Welt aus einer indefiniten Mannigfaltigkeit von Informationseinheiten – also Bits – besteht, dann müssen auch Entscheidungen bzw. Kontexte anerkannt werden, die angeben, welche Informationen, welche Bits für bestimmte Berechnungen wichtig sind und welche nicht. Wird dies akzeptiert, dann besteht die Welt eben nicht mehr homogen nur aus Informationen, sondern auch aus Kontexten von Informationen. Dies steht im Widerspruch zur Ausgangsannahme. Wird jedoch nun der Kontext selbst zur Information erklärt, so entsteht der Zirkelbezug, dass alles, was eine Information bestimmen soll, selbst wieder Information ist. Dies ist aber das Problem des Verhältnisses von Information und Bedeutung. Nach dem klassischen Paradigma wird nun Bedeutung auf Information reduziert, wodurch eine Binnenstruktur der Kontexte in sich zusammen fällt. Wiederum Kaehr wies deutlich darauf hin, dass jedoch „das transklassische Paradigma … auf der Irreduzibilität von Information und Bedeutung besteht." (Kaehr 1989) D.h. es gibt eine nichtklassische Forderung der Aufrechterhaltung der Kontexte, wonach das Eine nicht auf das Andere zurückführbar ist. Soweit zum Verhältnis von Transklassik und Digitalismus.

Dreyfus, Hubert L.; *Cybernetics as the last State of Metaphysics*. Akten des XVI. Int. Kongr. f. Philos. Wien, 1968, Band II, p. 493-499

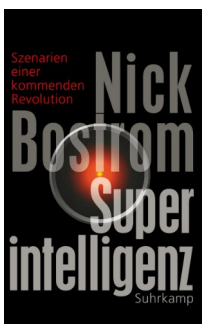
Dr. Joachim Paul, Jan. 2022

Nick Bostrom und die Seed-KI // AI

"Dies bringt uns zu einem weiteren wichtigen Konzept, dem der "rekursiven Selbstverbesserung". Eine erfolgreiche Seed-KI wäre in der Lage, sich selbst iterativ zu verbessern: Eine frühe Version der KI könnte eine verbesserte Version von sich selbst entwerfen, und die verbesserte Version - da sie intelligenter ist als das Original - könnte in der Lage sein, eine noch intelligentere Version von sich selbst zu entwickeln, und so weiter. [20] Unter bestimmten Bedingungen könnte ein solcher Prozess der rekursiven Selbstverbesserung lange genug andauern, um zu einer Intelligenzexplosion zu führen - ein Ereignis, bei dem die Intelligenz eines Systems in kurzer Zeit von einer relativ bescheidenen Ausstattung mit kognitiven Fähigkeiten (vielleicht in den meisten Bereichen untermenschlich, aber mit einem domänenspezifischen Talent für Programmierung und KI-Forschung) zu radikaler Superintelligenz ansteigt. Wir werden auf diese wichtige Möglichkeit in Kapitel 4 zurückkommen, wo die Dynamik eines solchen Ereignisses genauer analysiert wird. Man beachte, dass dieses Modell die Möglichkeit von Überraschungen nahelegt: Versuche, eine künstliche allgemeine Intelligenz zu schaffen, könnten so gut wie vollständig scheitern, bis die letzte fehlende kritische Komponente eingebaut ist, woraufhin eine Keimzelle der KI zu nachhaltiger rekursiver Selbstverbesserung fähig werden könnte."

"[20] Dies ist das von Good (1965) und Yudkowsky (2007) beschriebene Szenario. Man könnte jedoch auch eine Alternative in Betracht ziehen, bei der die iterative Abfolge einige Schritte umfasst, die keine Intelligenz-verbesserung, sondern eine Vereinfachung des Designs beinhalten. Das heißt, in einigen Phasen könnte sich die Seed-KI selbst umschreiben, damit spätere Verbesserungen leichter zu finden sind."

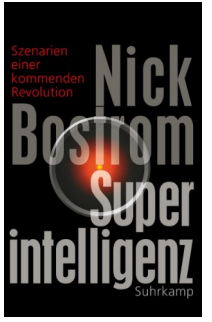
Nick Bostrom, *Superintelligenz – Szenarien einer kommenden Revolution*, Berlin 2014, S.50



Nick Bostroms Fußnoten

"Ein lautstarker Kritiker aus dieser Zeit war Hubert Dreyfus. Andere prominente Skeptiker aus dieser Zeit sind John Lucas, Roger Penrose und John Searle. Von diesen war jedoch nur Dreyfus hauptsächlich damit beschäftigt, Behauptungen darüber zu widerlegen, welche praktischen Leistungen wir von den bestehenden Paradigmen in der KI erwarten sollten (obwohl er anscheinend offen für die Möglichkeit war, dass neue Paradigmen weiter gehen könnten). Searles Ziel waren funktionalistische Theorien in der Philosophie des Geistes, nicht die instrumentellen Fähigkeiten von KI-Systemen. Lucas und Penrose bestritten, dass ein klassischer Computer jemals so programmiert werden könnte, dass er alles tun kann, was ein menschlicher Mathematiker tun kann, aber sie bestritten nicht, dass irgendeine bestimmte Funktion prinzipiell automatisiert werden könnte oder dass KI-Systeme schließlich sehr leistungsfähig werden könnten. Cicero bemerkte, dass "nichts so absurd ist, als dass nicht irgendein Philosoph es gesagt hätte" (Cicero 1923, 119); dennoch ist es erstaunlich schwer, sich an einen bedeutenden Denker zu erinnern, der die Möglichkeit einer maschinellen Superintelligenz in dem in diesem Buch verwendeten Sinne bestritten hat."

Nick Bostrom, *Superintelligenz – Szenarien einer kommenden Revolution*, Berlin 2014, S.369



Stuart Russell - Human Compatible

"Schon vor der Geburt der KI im Jahr 1956 rümpften abgehobene Intellektuelle die Nase und behaupteten, es könne gar keine intelligenten Maschinen geben. Alan Turing widmete 1950 einen Großteil seiner wegweisenden Abhandlung »Computing Machinery and Intelligence« der Widerlegung dieser Behauptungen. Doch noch immer muss sich die KI-Community fortwährend ähnlicher Aussagen erwehren. Zu den Verfechtern einer Unmöglichkeit der KI gehören Philosophen [6], Mathematiker [7] und andere. In der aktuellen Debatte zur Superintelligenz haben mehrere Philosophen diese Behauptungen wieder aufgegriffen, um zu beweisen, dass die Menschheit sich nicht fürchten muss.[8,9] Das kommt wenig überraschend."

[6] Ein wichtiger früher Beitrag, in dem die Aussichten regelbasierter KI-Systeme infrage gestellt werden: Hubert Dreyfus, Was Computer nicht können. Die Grenzen künstlicher Intelligenz (1989).

[7] Das erste einer Reihe von Büchern, in denen physikalische Erklärungen für das Bewusstsein gesucht und Zweifel über die Fähigkeit von KI-Systemen, jemals echte Intelligenz zu erreichen, geäußert werden: Roger Penrose, Computerdenken (1991, 2009).

[8] Eine Neuauflage der KI-Kritik auf Grundlage des Unvollständigkeitssatzes: Luciano Floridi, »Should we be afraid of AI?« Aeon, 9. Mai 2016.

[9] Eine Neuauflage der KI-Kritik auf Grundlage des Arguments vom Chinesischen Zimmer: John Searle, »What your computer can't know«, The New York Review of Books, 9. Oktober 2014.

Stuart Russell – *Human Compatible*, Frechen 2019, Longlisted for the 2019 Financial Times and McKinsey Business Book of the Year Award, W Econ. Forum

Stuart Russell - Human Compatible, die 2te

"Die »One Hundred Year Study on Artificial Intelligence«, kurz AI100, ist ein ambitioniertes Langzeitprojekt der Stanford University. Sie soll den Fortschritt der KI im Auge behalten oder – wie die Autoren es ausdrücken – »untersuchen und antizipieren, wie sich die künstliche Intelligenz auf alle Aspekte der Arbeit, des Lebens und des Spielens der Menschen auswirken wird«. Der erste große Zwischenbericht mit dem Titel »Artificial Intelligence and Life in 2030« (Künstliche Intelligenz und das Leben im Jahr 2030) hält jedoch eine Überraschung bereit:[10] Wie vielleicht zu erwarten ist, betont er den Nutzen der KI in Gebieten wie der medizinischen Diagnose und der Kraftfahrzeugsicherheit. Eher nicht zu erwarten ist die Behauptung, dass »eine Rasse übermenschlicher Roboter […] anders als in Kinofilmen nicht zu erwarten und vermutlich komplett unmöglich« ist.
Meines Wissens ist dies das erste Mal, dass ernsthafte KI-Forscher öffentlich die Ansicht vertreten, dass eine dem Menschen ebenbürtige oder gar übermenschliche KI ein Ding der Unmöglichkeit ist. Diese Aussage kommt noch dazu mitten in einer Phase extrem rascher Fortschritte in der KI-Forschung, in der eine Grenze nach der anderen durchbrochen wird."

[10] Ein Bericht verdienter KI-Forscher, die behaupten, dass eine übermenschliche KI voraussichtlich unmöglich ist: Stone, Peter et al., »Artificial intelligence and life in 2030«, One Hundred Year Study on Artificial Intelligence, Bericht des Studienpodiums 2015 aus dem Jahr 2016.

Stuart Russell – *Human Compatible*, Frechen 2019

Jürgen Geuter, 6 Digitalisierungsmythen

"Irrtum 1: Die Anwendung von Ethik kann man in Computerprogrammen formulieren.

Was ist eigentlich Intelligenz?
Irrtum 2: Daten erzeugen Wahrheit, und falls nicht, braucht man einfach mehr Daten."

Irrtum 3: In 20 Jahren gibt es eine künstliche Intelligenz, die genauso gut wie oder besser ist als menschliche.

Gesetze als Code? Funktionieren nicht
Irrtum 4: Diskriminierung durch Algorithmen ist schlimmer als Diskriminierung durch Menschen.

Irrtum 5: Gesetze und Verträge können in Code ausgedrückt werden, um ihre Anwendung zu standardisieren.

Open Source ist nicht gleich Freiheit
Irrtum 6: Digitale Freiheit drückt sich in der vollständigen Autonomie des Individuums aus."

Jürgen Geuter, *Nein, Ethik kann man nicht programmieren*, DIE ZEIT online, 27.11.2018

Zwischenfazit

Ankündigungsstatements der KI-Konstrukteure ...

Die dem Menschen überlegene, allgemeine Intelligenz,
die Superintelligenz, kommt in 20 bis spätestens 30,35 Jahren.

Und zwar seit der Dartmouth-Conference im Sommer 1956.

Einige Statements in Kurzform

Rudolf Kaehr 1993: *"Die heutige künstliche Intelligenz kommt deswegen nicht voran und bleibt vollständig in einem mechanistischen Denkbild verhaftet, weil sie eben die Subjektivität nicht in die Maschine mit einbezieht. […]*
… dann können wir nicht bei der Kognition haltmachen, weil wir dann kapieren müssen, dass es ... Lebewesen sind, die Kognitionen haben.
… dass der Konstrukteur, dann, wenn er die Maschine ... konstruiert hat, eben weiß, was er getan hat, und weiß, warum sie funktioniert und wie sie funktioniert. Also das heißt, er muss auch die Grenzen dieser Maschine kennen."

John Searle 2014: *"Die seltsame Verbindung von Behaviorismus - jedes System, das sich so verhält, als ob es einen Verstand hätte, hat auch wirklich einen - und Dualismus - der Geist ist kein gewöhnlicher Teil der physikalischen, biologischen Welt wie die Verdauung - hat zu den Verwirrungen geführt, die dringend aufgeklärt werden müssen."*

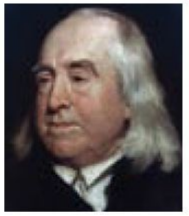
→ Lokalisierungsargument Kaehr, Memristor

NOTABLE MOMENTS IN MACHINE ETHICS

When Science Imitates Art

Long before ethicists, roboticists and AI experts became interested in the possible ethical ramifications of robots' behavior, science-fiction writers and film directors toyed with scenarios that were not always unrealistic. In recent years, however, machine ethics has become a bona fide field of research, in part drawing inspiration from the writings of 18th-century philosophers.

1495 Leonardo da Vinci designs one of the first humanoid robots



1780s Jeremy Bentham (*above*) and John Stuart Mill propose that ethics is computable

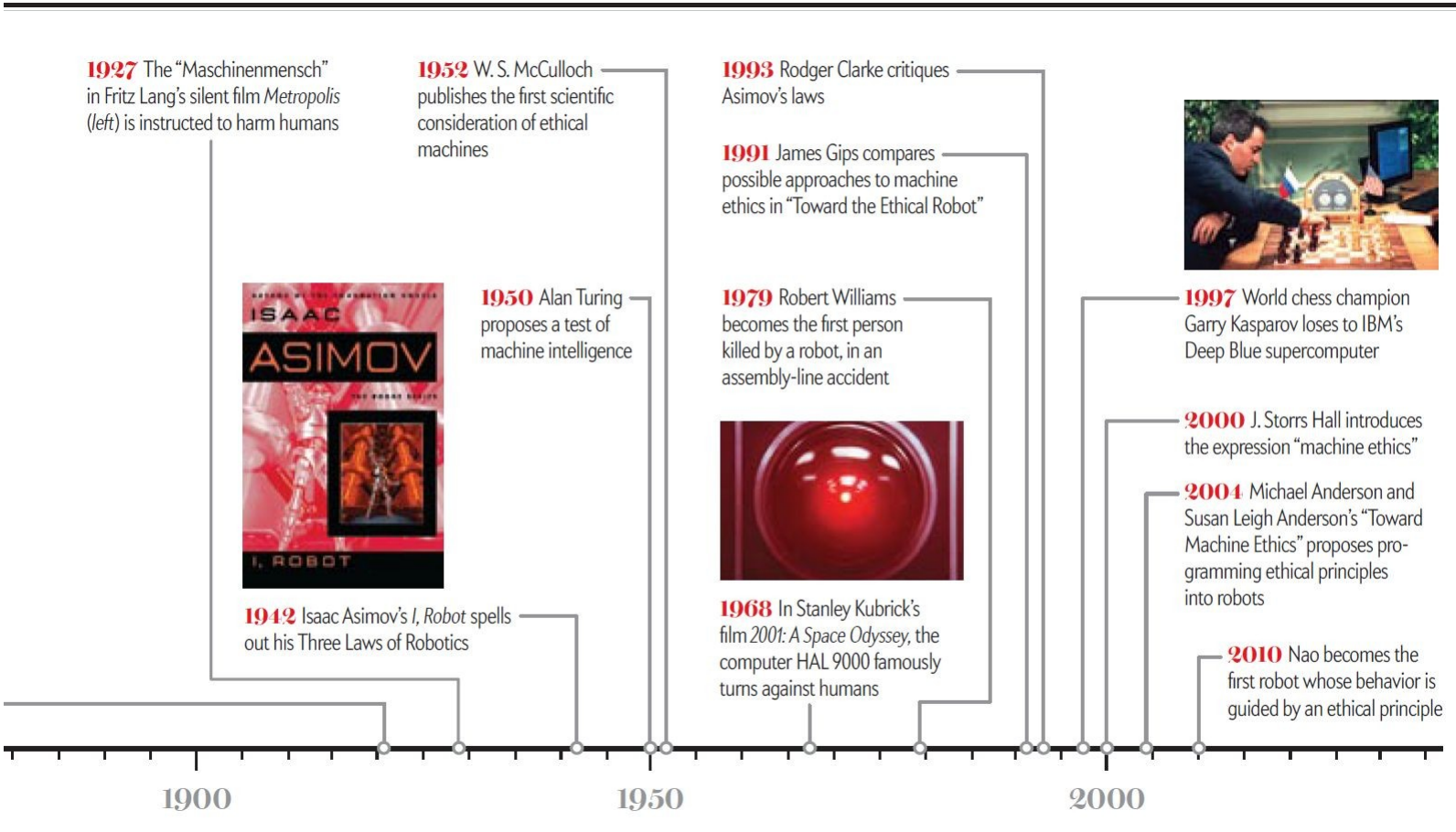
1921 Karel Čapek's play *R.U.R* first introduces the word "robot" and the concept of robot rebellion

1750 1800 1850

Maschinenethik
Zeitleiste nach
Anderson/
Anderson 2010

Teil 1

Aus
„Robot Be Good"
Scientific American



Maschinenethik
Zeitleiste nach
Anderson/
Anderson 2010

Teil 2

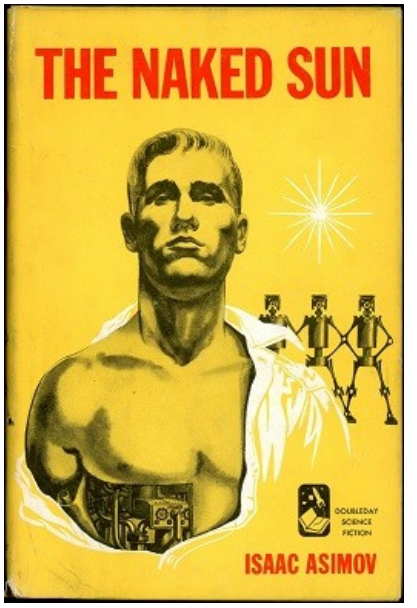
Aus
„Robot Be Good“
Scientific American

Isaac Asimovs Robotergesetze … ein alter Hut, aber ...

"1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law."

In Asimovs Foundation-Erzählkosmos entwickeln die beiden Roboter R. Giskard Reventlow und R. Daneel Olivaw via Kommunikation miteinander ein nulltes Gesetz:
"0. A robot may not harm humanity, or through inaction, allow humanity to come to harm."

Reventlow kostet die Anwendung seine Existenz.
Olivaw gelingt die allmähliche Anpassung.



Welche Arbeit von McCulloch meinen die Andersons?

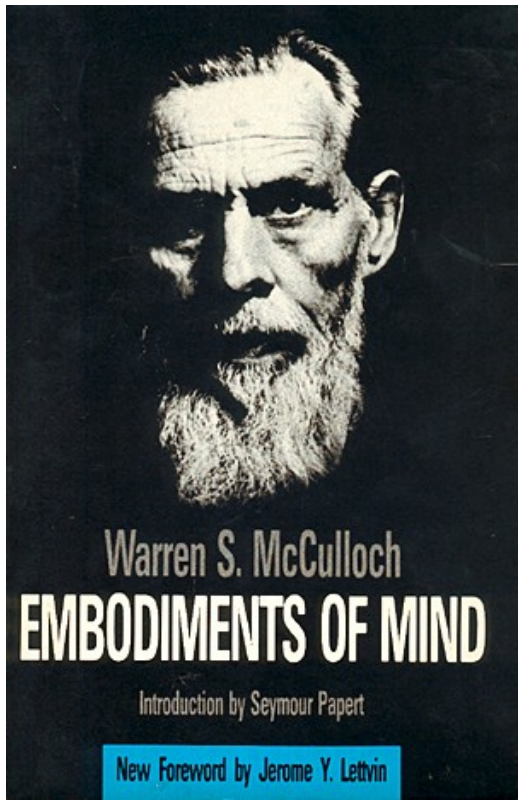
Toward Some Circuitry of Ethical Robots
or an Observational Science of the Genesis of Social Evaluation
in the Mind Like Behavior of Artifacts

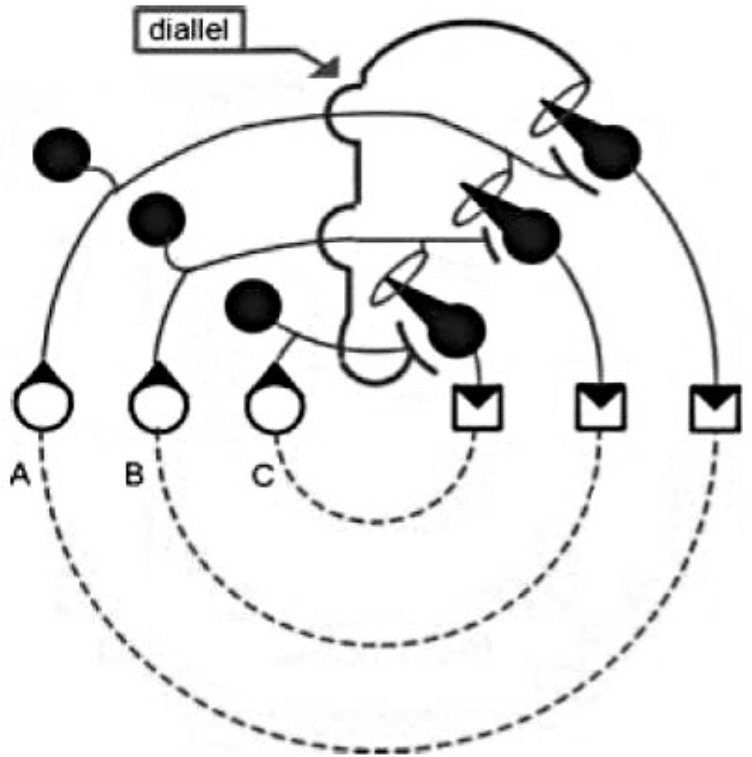
Zu Schaltkreisen ethischer Roboter oder: Eine
Beobachtungswissenschaft der Genese sozialer Bewertungen
im verstandesähnlichen Verhalten von Artefakten

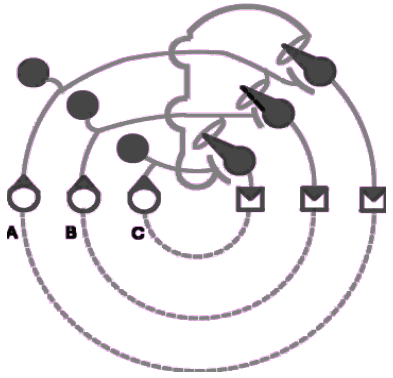
Warren S. McCulloch, Toward … Artifacts, Acta Biotheoretica, Vol XI, 1956, pp. 147-156,
reprinted in:
Warren S. McCulloch, Embodiments of Mind, Cambridge, Mass., 1970, pp. 194-202

New German translation by Joachim Paul (with the help of deepl.com)
online: www.vordenker.de Neuss 2019, J. Paul (Ed.), ISSN 1619-9324
URL des Beitrags: < https://www.vordenker.de/ggphilosophy/mcc_ethical_en_ger.pdf >

Heterarchisches Elementarnetz nach W.S. McCulloch







Zum topologischen! Argumentationsgang McCullochs:

- die strenge Präferenz- oder Vorzugsrelation verhält sich isomorph zur logischen Implikation,
ein Beweis lässt sich einfach über Wertetabellen erbringen.

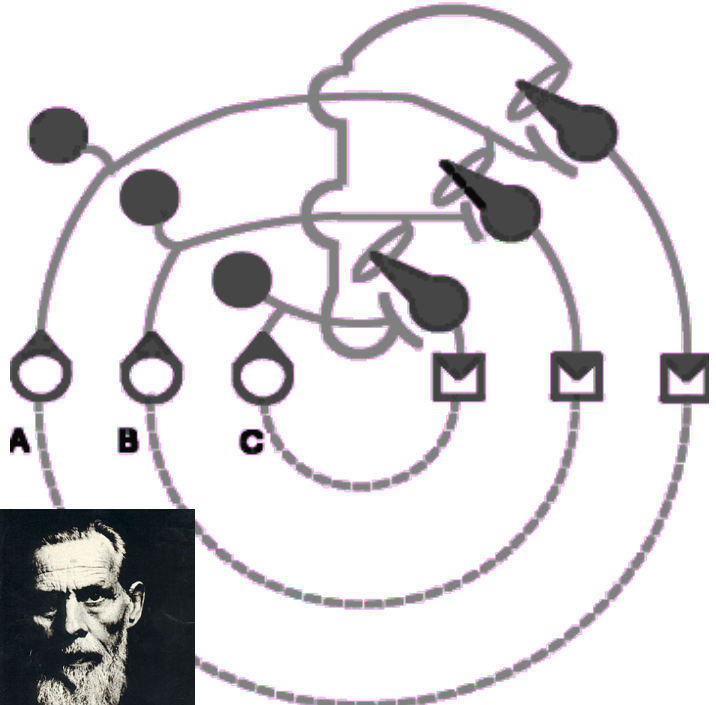
- das Transitivitätsgesetz der klassischen Logik:

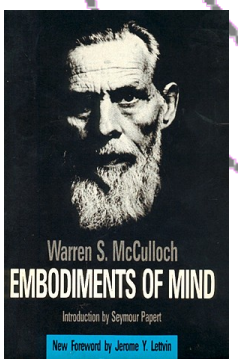
$$(A \rightarrow B) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C)$$

- wird durch das Verhalten des McCulloch-Netzes verletzt, da dieses - vom Standpunkt der klassischen Logik aus betrachtet – ganz offensichtlich Nonsens produziert:

$$(A \rightarrow B) \wedge (B \rightarrow C) \rightarrow (A \leftarrow C)$$

Heterarchisches Elementarnetz nach W.S. McCulloch





"... Consider the case of three choices, A or B, B or C, and A or C in which A is preferred to B, B to C, and C to A. The irreducible nervous net is shown in Figure 4. It requires one diallel in the plane. Its three heterodromic, branches link the dromes so as to form a circle in the net which is distinguished from an endrome in that it is not the circuit of any drome but transverse to all dromes, i.e., diadromic. The simplest surface on which this net maps topologically (without a diallel) is a tore. Circularities in preference instead of indicating inconsistencies, actually demonstrate consistency of a higher order than had been dreamed of in our philosophy. An organism possessed of this nervous system 🕮 six neurons 🕮 is sufficiently endowed to be unpredictable from any theory founded on a scale of values. It has a heterarchy of values, and is thus internectively too rich to submit to a summum bonum*..."*

Warren S. McCulloch, A Heterarchy of Values Determined by the Topology of Nervous Nets, Bull. Math. Biophys. 7 (1945) 89-93

Gregory Batesons categories of learning

"Lernen 0 ist durch die spezifische Wirksamkeit der Reaktion charakterisiert die - zu Recht oder zu Unrecht - keiner Korrektur unterliegt.
[fest verdrahtet, die M. IST der Algorithmus]

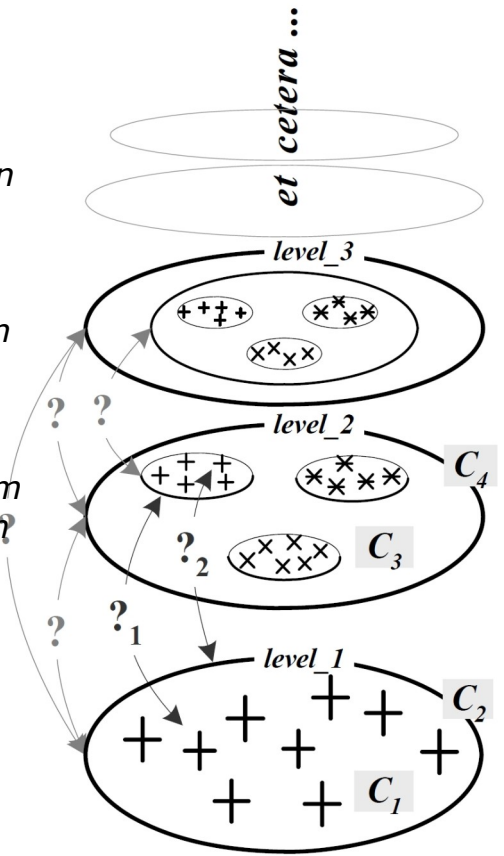
Lernen 1 ist Veränderung in der spezifischen Wirksamkeit der Reaktion durch Korrektur von Irrtümern der Auswahl innerhalb einer Menge von Alternativen.
[ANNs, Pavlov, Behaviorismus, bedingter Reflex]

Lernen 2 ist Veränderung im Prozess des Lernens 1, z.B. eine korrigierende Veränderung in der Menge von Alternativen, unter denen die Auswahl getroffen wird, oder es ist eine Veränderung in der Art und Weise wie die Abfolge der Erfahrung interpretiert wird.

Lernen 3 ist Veränderung im Prozess des Lernens 2, z.B. eine korrigierende Veränderung im System der Mengen von Alternativen, unter denen die Auswahl getroffen wird. (Wir werden später sehen, dass es manchmal pathogen ist, diese Leistungsstufe von einigen Menschen und einigen Säugetieren zu verlangen.)

Lernen 4 wäre Veränderung im Lernen 3, kommt aber vermutlich bei keinem ausgewachsenen lebenden Organismus auf dieser Erde vor. Der Evolutionsprozess hat jedoch Organismen hervorgebracht, deren Ontogenese sie zum Lernen 3 bringt. Die Verbindung von Ontogenese und Phylogenese erreicht in der Tat Ebene 4."

Gregory Bateson, *Ökologie des Geistes*, Frankfurt a.M. 1999, S.379



Psychologische Effekte … religiöse Rückbindungen

Schöpfungsanalogie:

Gott → Mensch → Robot (Kommentar Peter Sloterdijk aus Essay *Götterdämmerung*)

Siehe u.a. Homunkulus-Tradition, alchemistische Tradition des Denkens (G. Günther)

Sünden, Schuld und Strafen:
Monotheismen: Jüdisch, christlich, muslimisch: Sündenfall aus Genesis, treben nach Erkenntnis, Baum der Erkenntnis
Polytheismen:
Altgriechisch → Prometheus und Ikaros
Babylonisch → Turmbau, Gilgamesch und Enkidu (Gilgamesch-Epos)
Altpersisch, avestisch → Zarathustra, Ahura Mazda vs. Angra Mainyu (Ahriman), Vorstellung des jüngsten Gerichts
Mythische und literarische Kontrollverluste mit nachfolgenden Katastrophen:
Golem des Rabbi Löw (Prag)
J.W.v.Goethe, *Zauberlehrling,* Homunculus im Faust II
Mary Shelley, *Frankenstein oder der moderne Prometheus*

KI als Prothetik ...

Rudolf Kaehr 1993: *"In dem Sinn läßt sich vielleicht als Abrundung sagen, daß die Vollendung des Systems Mensch – wenn ich's mal technisch sagen darf – gegeben ist, erstens dadurch, daß er sich mit seiner Technik, die ihn generiert, verwebt, verquickt …"*

Mark Buchanan 2008: *"Done properly, computer simulation represents a kind of "telescope for the mind," multiplying human powers of analysis and insight just as a telescope does our powers of vision. With simulations, we can discover relationships that the unaided human mind, or even the human mind aided with the best mathematical analysis, would never grasp."*
Mark Buchanan, *This Economy Does Not Compute*, NYT 01.10.2008

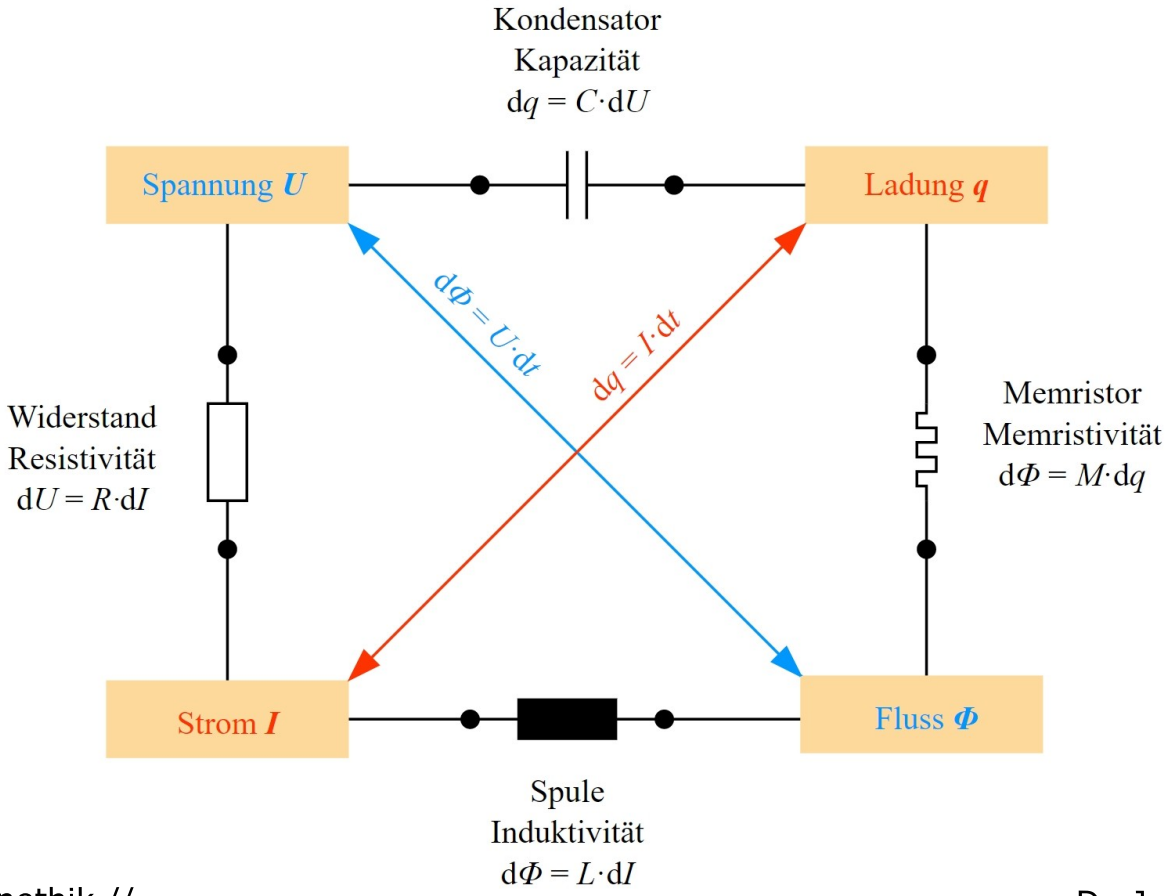
Heike Gfrereis 2013: *"Zettelkästen. Maschinen der Phantasie"*
(Buchtitel) Stichwort Serendipität

Thesen:
notwendige Bedingungen für die Realisation einer KI

- → nicht Turingsch, → nicht von Neumannsch
  (ANN-Prozessoren erfüllen Letzteres, TrueNorth (IBM), TensorFlow (Google), abber abbildbar auf Algorithmen)
- Echte, irreduzible Parallelität
- Hardwarebezug, Lokalisierbarkeit der Lernprozesse
- Einschreibungen in Materie // Hardware
- Keine heutigen Hardwarestrukturen – anders strukturierte ALUs in Prozessoren
- Polylogisch
- Polykontextural
- Echte Synchronizität
- Konstrukteure kennen Grenzen der Maschine

- Mindestens präzise Begriffsdefinitionen, notwendig, jedoch nicht hinreichend! Z.B. Busy Beaver TMs

Der Memristor – Einschreiben in Materie ...



Herzlichen Dank
für Ihre Aufmerksamkeit!

Anhang

Bildverweise

HAL 9000
Von Grafiker61 - Eigenes Werk, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=46424786

alphafold
Von Holger87 - Eigenes Werk, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=20396685

Myoglobin
Von →AzaToth - self made based on PDB entry, Gemeinfrei, https://coVon — MovGP0 - Eigenes Werk, Copyrighted free use, https://commons.wikimedia.org/w/index.php?curid=51119087
Von Operarius - File:Stonehenge, Salisbury.JPG, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4493066

Von Ricardo Liberato - All Gizah Pyramids, CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=2258048

By May be found at the following website: http://www.libraries.wvu.edu/exhibits/asimov/rare/mton.htm ., Fair use, https://en.wikipedia.org/w/index.php?curid=20132101

Neuron
Von Chrislb - Erstellt von Chrislb, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=224561

Memristor
Von — MovGP0 - Eigenes Werk, Copyrighted free use, https://commons.wikimedia.org/w/index.php?curid=51119087