Publisher: IISI - International Institute for Socio-Informatics
ISSN 1861-4280

international reports
on socio-informatics

volume 20 issue 1
2023

»Künstliche Intelligenz«: Dichtung und Wahrheit

Einblicke in die Technik des Berechnens und in Mythen um Intelligenz

Peter Brödner

Honorarprofessor
Universität Siegen, Wirtschaftsinformatik

↑Inhalt↑

The 'international reports on socio-informatics' are an online report series of the International Institute for Socio-Informatics, Bonn, Germany. They aim to contribute to current research discourses in the fields of 'Human-Computer-Interaction' and 'Computers and Society'. The 'international reports on socio-informatics' appear at least two times per year and are exclusively published on the website of the IISI.

Impressum

IISI - International Institute for Socio-Informatics
Stiftsgasse 25
53111 Bonn
Germany

fon: +49 228 6910-43
mail: iisi@iisi.de
web: http://www.iisi.de

↑Inhalt↑

Vorwort

Im Hauptteil dieser Online-Publikation ist eine Reihe von Aufsätzen aus meiner Feder versammelt, die sich kritisch mit neueren, auf sog. ›subsymbolischen‹ bzw. ›konnektionistischen‹ Ansätzen beruhender »künstlicher Intelligenz« (»KI«) auseinandersetzen. Es handelt sich (mit Ausnahme des letzten) durchweg um Veröffentlichungen in internationalen wissenschaftlichen Fachzeitschriften und eingeladene Konferenz- oder Buchbeiträge, die hier in der Preprint-Fassung wiedergegeben werden.

Die Motivation zur gesammelten Reproduktion der Beiträge entstammt der Beobachtung, dass öffentliche Diskurse über »KI« weit über Fachzirkel hinaus erneut hohe Wellen schlagen, nachdem sie in den 1990er und 2000er Jahren fast ganz abgeebbt waren. Aus dieser Zeit stammt auch meine Buchpublikation »Der überlistete Odysseus« (Brödner 1997), die sich bereits kritisch mit dem konnektionistischen Paradigma der »KI« auseinandersetzte, aber im damaligen »KI-Winter« auf wenig Resonanz stieß. Heute werden in kurzen Abständen immer wieder neue auf diesem Ansatz beruhende, angebliche Erfolge gemeldet und als vermeintliche Durchbrüche gefeiert. Ein »KI«-Hype jagt in den letzten Jahren den nächsten. Umso notwendiger erscheint es, abermals zu reflektieren, inwieweit dieser jeweilige Überschwang gerechtfertigt ist. Schon wenige Blicke ›unter die Haube‹ der so gepriesenen Systeme, auf die dahinter stehenden Methoden und Funktionsweisen, lassen daran Zweifel aufkommen, denen in den Beiträgen nachgegangen wird.

Thematisch behandeln die Beiträge verschiedene Aspekte der neueren »KI«-Entwicklungen aus ganz unterschiedlichen Perspektiven. Zudem sind sie auf den jeweiligen Kontext ausgerichtet, in dem sie erschienen sind (s. die jeweiligen Hinweise dazu bei der Vorstellung der Beiträge). Darin unterscheiden sie sich voneinander. Zugleich müssen sie aber auch jeweils für sich allein verständlich sein und dazu auf gewisse gemeinsame Grundlagen Bezug nehmen. In dieser Hinsicht sind sie natürlich in einzelnen Passagen redundant. Zusammengenommen fügen sich die unterschiedlichen Perspektiven und die Aspekte, die sie in den Blick nehmen, zu einem einigermaßen kohärenten Bild, das diese Zusammenstellung lohnend erscheinen lässt.

Zur Einführung in die gemeinsamen Grundlagen ist den reproduzierten Beiträgen ein eigens verfasster Aufsatz vorangestellt, in dem übergreifende Aspekte behandelt und Zusammenhänge dargestellt werden. Damit wird Hintergrundwissen und eine Art kognitive Basis für die Lektüre der Einzelbeiträge bereitgestellt.

Karlsruhe, im März 2023

Peter Brödner

↑Inhalt↑

Inhalt

↑Inhalt↑

Ist Verstand mit Logik, Denken mit Berechnen gleichzusetzen?

1 Einführung: Berechnungsmethoden ›unter der Haube‹

Erst kürzlich hat der öffentliche Diskurs um die sog. »künstliche Intelligenz« mit dem Erscheinen der Sprachverarbeitungs-Software GPT-3 und ChatGPT der Fa. OpenAI (GPT steht für »generative pre-trained transformer«) erneut neue Wellen öffentlicher Erregung aufgeworfen – ein Durchbruch technischer Realisierung von ›Intelligenz‹ für die einen, eine Allmachtsfantasien entsprungene Illusion und Quelle großer gesellschaftlicher Risiken für die anderen. Mittels des Sprachmodells GPT lassen sich aus vorgegebenen Sprachtexten neue anschlussfähige Texte erzeugen, die in einer scheinbar bedeutungsvollen Beziehung zueinander stehen, sei es in Form von Übersetzungen, Frage-Antwort-Spielen oder der Erzeugung von Bildern aus Texten (Vaswani et al. 2017).

Am Beispiel GPT lassen sich wie in einem Kristall grundlegende funktionale Aspekte aufzeigen, die in Debatten um »KI« erstaunlicherweise kaum eine Rolle spielen, aber für die Einschätzung des Entwicklungsstandes und damit implizierter Potenziale und Risiken wesentlich sind. Wie viele auf »konnektionistischen« Ansätzen bzw. »maschinellem Lernen« beruhende »KI«-Methoden nutzen auch GPT »künstliche neuronale Netze (KNN)« mit einer sehr großen Zahl zu bestimmender Parameter (derzeit: $175 \cdot 10^9$) als funktionalem Kern. Solche »KNN« werden durch ihre Konstrukteure aus Erfahrung, aber durch theorieloses Probieren für eine Aufgabe strukturiert. Das erfordert bei Sprachmodellen eine hoch komplexe Struktur und extrem großen Aufwand für das ›Training‹ der Netze durch Berechnung der Verbindungsgewichte als Parametern. Mathematisch modellieren lassen sich »KNN« mittels Gleichungen linearer Algebra (Matrizenkalkül), woraus am Ende die Berechnungsverfahren und deren Programmierung hergeleitet werden (vgl. *Abb. 1*). Deren maschinelle Ausführbarkeit beruht letztendlich allein darauf, dass binäre elektronische Schaltsysteme die elementaren logischen Operationen physisch verkörpern, mittels derer sich auch arithmetische Operationen realisieren und programmgesteuert in getakteter Abfolge vollziehen lassen (vgl. hierzu die näheren Ausführungen in Kap. 3.2 des Beitrags »Das Produktivitätsparadoxon der Computertechnik«).

Anders als in Debatten häufig suggeriert, werden auch in »KI«-Systemen lediglich programmierte Berechnungsverfahren maschinell ausgeführt wie in jedem ›gewöhnlichen‹ Computersystem auch, nur gründen sie auf anderen mathem-

atischen Modellen und erfordern wegen hohen Rechenaufwands besonders leistungsfähige Hardware. Computer heißen so, weil sie komplizierte Berechnungen programmgesteuert durchführen – ohne jedes Geheimnis. Sie erzeugen aus vorgegebenen Eingabedaten kausal eindeutig determinierte Ausgabedaten – wie eine Funktion $f(x)$, die mit Eingabe des Arguments x stets einen bestimmten Funktionswert $f(x)$ erzeugt (egal wie kompliziert die Funktion f aufgebaut ist).

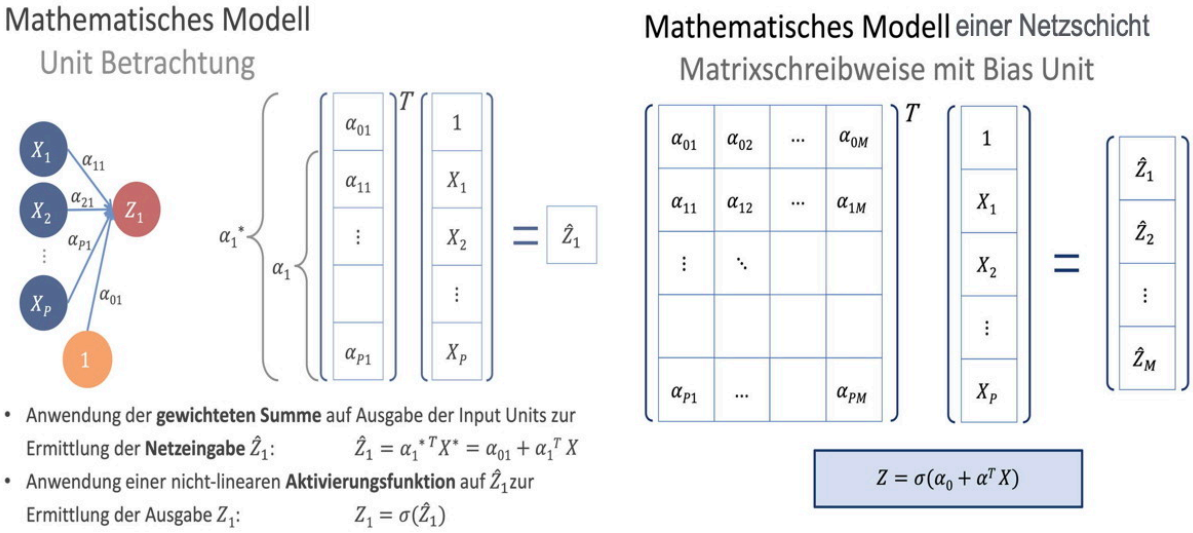


Abb. 1: Mathematische Modellierung von »KNN« (Darstellung: Uni. Köln)

Das Besondere beim Einsatz von »KNN« ist nun, dass zunächst nur die Struktur des Netzwerks für die jeweils zu lösende Aufgabe festgelegt wird und deren noch unbestimmte Parameter, die Elemente der Matrizen als Verbindungsgewichte, noch zu berechnen sind. Dies geschieht aufgabenspezifisch durch unterschiedliche Approximationsverfahren zur Anpassung der Netzwerk-Funktion an sehr große Mengen von außen vorgegebener Daten mittels verschiedener mathematischer Schätzmethoden (z.B. Minimierung einer Verlustfunktion, Maximum-Likelihood und dgl.) – ein aufwendiger Vorgang der als ›Training‹ des Netzwerks bezeichnet wird (aber nichts mit gewohntem Lernen zu tun hat). Erst die derart ›trainierten‹ Netzwerke lassen sich dann auf die eigentlichen Aufgaben ansetzen.

Die Leistung des Sprachmodells GPT beruht im Kern darauf, für maschinelle Sprachverarbeitung passend strukturierte »KNN« zu ›trainieren‹, indem aufgrund eines riesigen Korpus verfügbarer Texte (> 10^2 TB) durch Modellierung und Abstraktion von Bedeutung rein syntaktische quantitative Beziehungen zwischen den einzelnen Wörtern des verwendeten Vokabulars berechnet werden. Dazu müssen die Wörter und ihre Kontextbeziehungen zunächst durch Vektoren als mathematisch handhabbaren Objekten repräsentiert und modelliert werden. Mit so berechneten bedingten Wahrscheinlichkeiten für Beziehungen zwischen Wörtern eines Kontexts wird bestimmt, welche Worte auch im zu generierenden Text höchst wahrscheinlich auf ein bereits gegebenes Wort folgen; so werden zur Genese neuer Texte durch Berechnung scheinbar ›vernünftig‹ fortgesetzte Wortfolgen ge-

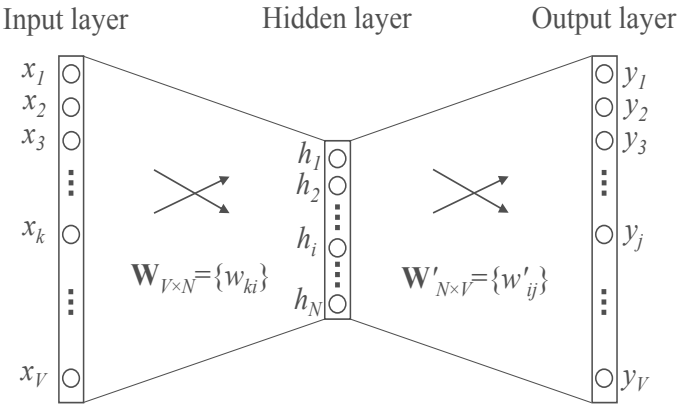
bildet, die freilich infolge dieser Modellierung und Abstraktion ihrer Bedeutung beraubt sind.

Aus diesem grundlegenden Zusammenhang ergibt sich bereits die Erklärung, warum diese Computersysteme sowohl verblüffende Ergebnisse menschenähnlich erscheinender Texte zustande bringen können als auch für Fehlleistungen prädestiniert sind, ohne dass Betrachter freilich imstande wären, beides auf Anhieb zu unterscheiden, wie dies Experimente mit GPT immer wieder demonstrieren. So haben paradoxerweise sowohl die Enthusiasten als auch die Skeptiker beide zugleich recht: Weil die berechneten bedingten Wahrscheinlichkeiten allein auf abstrakten quantitativen Beziehungen zwischen Wörtern allgemeinen Sprachgebrauchs beruhen und daher deren Bedeutung keine Rolle spielt, können sich gleichermaßen sprachlich gelungen erscheinende wie auch offensichtlich misslungene Resultate ergeben. Obgleich viele Texte syntaktisch und idiomatisch korrekt erscheinen, mithin Bedeutung vortäuschen, ist ihr Inhalt fragwürdig und nicht verlässlich. Infolge der Abstraktion fehlt den Programmen eben die Bedeutung der Wörter und sie ›verstehen‹ nicht, wovon die Rede ist. Ihnen fehlt der sinnhafte Bezug zur Welt. Zurecht werden sie daher auch als »*stochastische Papageien*« bezeichnet (Bender et al. 2021).

Dies lässt sich beispielhaft und prinzipiell anhand nachstehender Erläuterung der Funktionsweise eines KNN für Sprachverarbeitung aufzeigen, das auf einem einfachen Ein-Wort-Kontext-Modell (CBOW) beruht. Dieses übliche Verfahren des sog. »Word2Vector Embedding« wird oft auch als Eingangskomponente von komplexeren Modellen maschineller Sprachverarbeitung genutzt. Damit lassen sich Wörter und ihre Beziehungen zum Kontext mittels Vektoren quantitativ repräsentieren und ggf. in weiteren KNN-Komponenten (etwa in »En-» und »Decodern« von ChatGPT oder Übersetzungsmodellen) weiter verarbeiten (vgl. Xin 2016):

Mit V als Vokabulargröße (Anzahl Worte w_j, z.B. 1.000) und N (N < V) als Größe der hidden layer **h** wird diese aus der Eingabeschicht gebildet, indem mit $x_k = 1$ (k-tes Wort) und $x_j = 0$ für $j \neq k$ in der hidden layer jedes Wort w_k desVokabulars durch die k-te Zeile von **W** als Vektor repräsentiert wird:

$$\mathbf{h} = \mathbf{W}^T\mathbf{x} = \mathbf{W}^T_{(k,\cdot)} := \mathbf{v}^T_{w_I}$$ (Wort-Eingabevektor; vgl. nachstehende Grafik).



Simples Ein-Wort-Kontext-Modell (CBOW)

Mit der Matrix **W**' = {w'$_{ij}$} wird zudem für jedes Wort w$_j$ ein Zahlenwert u_j berechnet:

$$u_j = \mathbf{v'}_{w_j}{}^T \mathbf{h} \qquad \text{wobei } \mathbf{v}\text{'}_{wj} \text{ der Wort-Ausgabevektor ist.}$$

Damit werden in der Ausgabeschicht für jedes Wort w$_j$ unter Verwendung der sog. Softmax-Funktion (als einem üblichen log-linearen Klassifikationsmodell) die bedingten Wahrscheinlichkeiten berechnet:

$$p(w_j|w_I) = y_j = \frac{\exp(u_j)}{\sum_{j'=1}^{V} \exp(u_{j'})} \quad \text{bzw.} \quad p(w_j|w_I) = \frac{\exp\left(\mathbf{v'}_{w_j}{}^T \mathbf{v}_{w_I}\right)}{\sum_{j'=1}^{V} \exp\left(\mathbf{v'}_{w_{j'}}{}^T \mathbf{v}_{w_I}\right)}$$

Aus der Forderung nach Maximierung dieser bedingten Wahrscheinlichkeiten für alle j ∈ {1, …,V} – eine sehr aufwendige Berechnung – ergibt sich mit der Verlustfunktion $E = -\log p(w_O|w_I)$ und der Lernrate η die Korrektur der Gewichte w'$_{ij}$ zwischen hidden und output layer zu:

$$\mathbf{v'}_{w_j}{}^{(\text{new})} = \mathbf{v'}_{w_j}{}^{(\text{old})} - \eta \cdot e_j \cdot \mathbf{h}$$

sowie der w$_{ki}$ zwischen input und hidden layer zu:

$$\mathbf{v}_{w_I}^{(\text{new})} = \mathbf{v}_{w_I}^{(\text{old})} - \eta \mathrm{EH}^T$$

Der *N*-dimensionale Vektor EH^T (als gewichtete Summe der Ausgabevektoren der Worte w_k des Vokabulars) fügt dem Eingabevektor des Kontext-Worts gewisse Anteile jedes Ausgabevektors des Vokabulars hinzu. Damit bildet er abstrakte, quantitativ gewichtete, quasi-›semantische‹ Wortbeziehungen ab. Die damit generierten Wortfolgen sind zwar syntaktisch korrekt, täuschen damit Bedeutung aber nur vor. Das Verfahren kann zwecks Verbesserung der Leistung auch noch Nachbarwörter einbeziehen, steht oft am Anfang der Sprachverarbeitung und wird dann mit weiteren KNN-Modulen kombiniert (Vaswani et al. 2017).

2 »KI«-Entwicklung: Stets wiederkehrende Muster

Der Hype um die auf dem Modell GPT beruhenden Systeme zur Sprachverarbeitung ist in vielfacher Hinsicht besonders aufschlussreich und weist im historischen Rückblick – in Gestalt wiederkehrender Muster – eine Reihe typischer Merkmale der allgemeinen »KI«-Entwicklung auf.

Erstens entspricht auch der Hype um das Sprachmodell GPT als neuestem Coup der als »KI« bezeichneten Computertechnik einmal mehr einem durchgängigen Verlaufsmuster von zunächst euphorischem Überschwang, bald gefolgt von tiefer Ernüchterung, der nach genauerer Betrachtung verlangt. Dieses wellenförmige Muster tritt seit den ersten Tagen der physischen Realisierung programmierbarer Computer in den späten 1940er Jahren in Erscheinung (wobei aber die

entscheidenden ideellen Vorarbeiten deutlich älter sind; zum Wellenverlauf vgl. auch Teich 2020). Damals war bereits verbreitet und euphorisch von »Elektronen-Gehirnen« und »denkenden Maschinen« (Berkeley 1949) die Rede, die vermeintlich »schneller denken als Einstein« (Philadelphia Evening Bulletin 15.02.1946). Diese erste Welle verlief sich wieder, als man sich der im Vergleich zu heute verfügbarer Rechenleistung kläglichen Leistungsfähigkeit der Hardware bewusst wurde und fortan nüchterner – und zutreffender Weise – lieber von »Rechentechnik« oder »elektronischer Datenverarbeitung« sprach.

Mitte der 1950er Jahre hat dann der amerikanische Mathematiker McCarthy mit anderen Pionieren der Computertechnik (darunter Marvin Minsky und Claude Shannon) einen Antrag an die Rockefeller-Foundation zur Förderung einer Sommer-Forschungsgruppe über »Artificial Intelligence (AI)« gestellt. Mit dieser Bezeichnung ihres Vorhabens setzten sie einen bis heute wirksamen Anspruch in die Welt und lieferten auch gleich eine Art Definition mit, die sich (wie oftmals auch heute noch) auf menschliches Verhalten bezieht: *»... making a machine behave in ways that would be called intelligent if a human were so behaving«* (McCarthy et al. 1955). Es ist gerade diese vage Unbestimmtheit des Anspruchs, die seine anhaltende Wirkung begründete, ihn zu einem wirkmächtigen, zweifellos genialen PR-Coup machte und im Laufe der Zeit viel Geld für wenig Ertrag strömen ließ. Greifbare Resultate hat das Projekt damals nicht erbracht, abgesehen von der durch den Ideen-Austausch möglicherweise inspirierten, Jahre später von McCarthy (1960) publizierten funktionalen Programmiersprache LISP (List Processor – tatsächlich ein großer Wurf). Danach gab es mehrere aufeinander folgende kleinere Wellen vermeintlicher Erfolge, jeweils gefolgt von anschließender Ernüchterung wie sie etwa von dem Philosophen Hubert Dreyfus im einzelnen beschrieben und grundlegend kritisiert wurde (Dreyfus 1979).

Das tragende Paradigma hinter diesen Versuchen ist die sog. »physical symbol hypothesis« (Newell 1980), die Annahme, ›Intelligenz‹ ließe sich letztlich rein formal durch komplexe logische Operationen auf einer Menge physisch repräsentierter, aber bedeutungsloser Symbole zustande bringen (weshalb dieser paradigmatische Zugang auch »symbolische KI« genannt wird). Diese in kleineren Wellen unterschiedlicher regelbasierter Ansätze und Versuche ablaufende Entwicklung kulminierte dann in den 1980er Jahren, auch vor dem Hintergrund wachsender Leistungsfähigkeit der Hardware, in Bemühungen um Realisierung groß angelegter »wissensbasierter Systeme« und »Expertensysteme«. Diese mächtig angeschwollene Welle zerschlug sich letztlich aber an der enttäuschenden, gleichwohl wachsenden Einsicht, dass menschliche Erfahrung und Kompetenz wesentlich auf implizitem, leiblich gebundenem Können beruht, das sich nur partiell, keineswegs vollständig in explizites Wissen transformieren und folglich auch nicht auf Logik reduzieren lässt (worauf schon Dreyfus 1979 u.v.a. stets hinwiesen; im Beitrag »Coping with Descartes' Error in Information Systems« wird genauer auf die Grenzen der Explikation von Können und Erfahrung in begrif-

fliches und propositionales Wissen eingegangen). Fortan herrschte längerer Zeit ein »KI-Winter« nur sehr geringer Entwicklungsaktivität.

Mit »künstlichen neuronalen Netzen« (KNN) gewann dann ein ganz anderer, auch sehr alter, aber lange mit einem Bann belegter Ansatz wieder allmählich wachsenden Zulauf, der auf dem Paradigma konnektionistischer bzw. »subsymbolischer KI« beruht. Kern dieses Ansatzes ist, wie am Beispiel von GPT gezeigt, rein erfahrungsbasiert, ohne theoretisches Fundament, Netzwerke ›künstlicher Neuronen‹ (McCulloch & Pitts 1943) für die jeweilige Aufgabe mutmaßlich passend zu strukturieren. Das Verhalten von KNN lässt sich mittels des Matrizenkalküls durch mathematische Funktionen mit einer großen Zahl von Parametern beschreiben (vgl. *Abb. 1*), damit auch für den Ablauf auf einem Computer programmieren; die Parameter werden dann durch programmierte Verfahren der Anpassung an große Mengen vorgegebener Daten berechnet. Das erfordert meist enormen Rechenaufwand, der inzwischen aufgrund exponentiellen Wachstums der Rechenleistung der Hardware meist bewältigt werden kann. Die eingangs erwähnten GPT-Systeme bilden eine Art Speerspitze dieser Entwicklung.

In Anbetracht dieses typischen Wellenmusters der »KI«-Entwicklung sprechen Sozialwissenschaftler auch mit gewisser Berechtigung von »promising technology« (Hirsch-Kreinsen 2023). Diesem Konzept zufolge werden verschiedene methodische Ansätze und ihre technische Umsetzung mit jeweils weitreichenden Versprechungen verknüpft, die oftmals zwar partiell, aber auf Dauer nicht wirklich eingelöst werden. So wird von Beginn an stets nach dem gleichen Muster argumentiert, man habe zwar vorläufig nur einen noch unvollkommenen Entwicklungsstand erreicht, der aber mit künftigen Forschungsanstrengungen zu durchschlagendem Erfolg geführt werden könne, wobei freilich die Möglichkeit, einen grundsätzlich verfehlten Ansatz zu verfolgen, stets außer Betracht bleibt. Damit lässt sich das ständige Hin-und-Her zwischen Überschwang und Ernüchterung recht plausibel erklären: Solange die Selbsttäuschung oder Illusion, die Denken mit Berechnen gleichsetzen zu können meint, sich durch oftmals verblüffende Ergebnisse neu bestätigt sieht, bestehen Anreize, weitere Ressourcen in die Entwicklung zu investieren, und je stärker sich die folgende Enttäuschung und Ernüchterung verbreiten, desto weniger neue Nahrung erhalten sie. Erst mit anhaltender wirklicher Ent-Täuschung ließe sich das Muster brechen.

Zweitens verweist der gegenwärtige Hype um GPT besonders deutlich auf die sich durch die ganze wellenförmige Entwicklung ziehende Diskrepanz zwischen Anspruch und Wirklichkeit, zwischen behaupteten Fähigkeiten und tatsächlicher Funktionsweise der technischen Systeme. Indem ständig versucht wird, nicht näher durchschaute Berechnungsverfahren mithilfe durchweg irreführend anthropomorphisierender Metaphern zu erklären, werden dafür unsachgemäße Vorstellungen von »Intelligenz« eingeschmuggelt. So werden rechnenden Maschinen menschliche Fähigkeiten zugeschrieben, die deren Funktionsweise aber gar nicht hervorzubringen imstande sind. Auf diesem Wege entstehen viele Missverständ-

nisse und vor allem weitreichende Illusionen über deren Leistungsfähigkeit: Beispielsweise mutieren Verfahren »maschinellen Lernens (ML)« unversehens in »lernende Maschinen«, fremdbestimmt programmgesteuerte, automatische (d.h. selbsttätige, aber fremdbestimmte, nicht *von selbst* tätige) Verfahrensabläufe avancieren zu »autonomen Systemen«, ein Artefakt wie GPT wird zu einer »KI«, die vermeintlich »Sprache versteht« oder aus sorgfältigen Problemanalysen entstammenden Berechnungsverfahren wird »künstliche Intelligenz« angedichtet. Zwar sind Rechnen und methodisches Berechnen zweifellos kognitive Tätigkeiten, Kognition und Denken sind aber nicht mit Berechnen gleichzusetzen, sondern gehen weit darüber hinaus, etwa wenn zur Analyse noch nicht durchschauter Probleme sowie zu deren anschließender Formalisierung und Modellierung Begriffsbildung, Intuition oder Kreativität gefordert sind (vgl. dazu den Beitrag *»Das Produktivitätsparadoxon der Computertechnik«* sowie weiter unten die Ausführungen zum Gödelschen Unvollständigkeitssatz).

Diese reduktionistische, letztlich auf schweren Kategorienfehlern beruhende Perspektive ignoriert die fundamentalen Unterschiede von implizitem Können und explizitem Wissen, von situativem Beurteilen und abstraktem Berechnen, von kreativer bzw. intuitiver »Abduktion« (Peirce 1934) und logischer Deduktion sowie deren Zusammenspiel im wirklichen Handeln lebendiger Personen bei der realen Bewältigung praktischer Probleme (diese basale Differenz betont bereits Weizenbaum 1977). Während der menschliche Geist stets über beide Fähigkeiten, über Urteilskraft und Berechnungskompetenz verfügt, über die Fähigkeiten, reale Vorgänge zu beurteilen und durch Berechnungsverfahren zu beschreiben, bleibt Computertechnik auf die Ausführung berechenbarer Funktionen beschränkt, wobei sich bereits relativ einfache Aufgaben als nicht berechenbar erweisen, z.B. das Halteproblem, die funktionale Äquivalenz zweier Programme oder der Nachweis der Widerspruchsfreiheit eines formalen Systems. Computer ›denken‹ nicht, sondern führen nur maschinell von Menschen bereits Gedachtes in Form von Algorithmen aus. Tatsächlich erweist sich »KI« als Bezeichnung für die maschinelle Ausführung zweckgemäß ausgedachter Berechnungsverfahren schlicht als Etikettenschwindel, mit dem die dazu benötigte kognitive Leistung irreführend den Berechnungsverfahren statt deren Schöpfern zugeschrieben wird.

Drittens hat es in wiederholt aufflammenden »KI«-Diskursen der mathematische Fachterminus »Algorithmus« zu erstaunlicher Prominenz gebracht, dessen laienhafter Gebrauch freilich seine eigentliche Bedeutung gänzlich verkennt und ins Reich der Mythen führt, wenn etwa von der »Macht der Algorithmen« oder gar von »lernenden Algorithmen« die Rede ist (einem Widerspruch in sich) oder davon, dass Algorithmen vermeintlich »intelligent« sind, »Bilder erkennen«, »Entscheidungen treffen«, »Sprache verstehen« etc.. Ein *Algorithmus* ist ein rein ideelles mathematisches Objekt, ein präzise formalisiertes Berechnungsverfahren, das aus Eingabedaten in endlich vielen Schritten ein Ergebnis erzeugt und anhält. Das Berechnungsverfahren kann auf verschiedene Weise in formaler Notation

dargestellt werden, z.B. als spezielle ›Turingmaschine‹ (Turing 1936) oder äquivalent als ein meist komplexer Funktions-Ausdruck des Lambda-Kalküls (Church 1941; Kleene 1952), der durch Funktionsanwendung (›β-Konversion‹) ausgewertet werden kann. Als mathematisches Objekt kann der Algorithmus auch durch ein *Programm* in einer ebenfalls formal spezifizierten Programmiersprache genau beschrieben werden, das dann, mittels eines Compilers automatisch in Maschinencode physischer Binärsignale übersetzt, auf einem binären Schaltsystem aus elementaren Logik- oder Speicherbausteinen ausführbar ist. Die symbolische Maschine ›Computer‹ führt mittels dieser kausal operierenden »Logikgatter«, die logische Funktionen physisch verkörpern, die per Programm formal beschriebenen berechenbaren Funktionen aus (ganz ohne zu ›wissen‹, was sie da ›tut‹ und wozu).

Eingedenk dessen wird ein nüchterner, die tatsächlichen Vorgänge und Funktionsweise in Betracht ziehender Blick benötigt, um sich angesichts der vielen vermenschlichenden und irreführenden Metaphern zur Beschreibung von »KI-Systemen« vor fehlgeleiteten Wahrnehmungen, Fehleinschätzungen möglicher Folgen und Fehlallokationen gesellschaftlicher Ressourcen zu schützen. Statt sich von immer wieder neu von Ergebnissen vermeintlich »intelligenter« technischer Funktionalität verblüffen zu lassen, wäre es weit sinnvoller und produktiver, nach Wegen zu suchen, auf denen mittels geeigneter Berechnungsverfahren kooperative kognitive Arbeit wirkungsvoll unterstützt und leistungsfähiger organisiert werden kann.

Viertens offenbart sich in der wellenförmigen Entwicklung der als »KI« apostrophierten Artefakte der Computertechnik auch ein tiefes Missverständnis dessen, was Technik und die Funktionsweise ihrer Artefakte im Allgemeinen und der Computertechnik im Besonderen ausmacht. Technik ist, wie schon Aristoteles (in der Nikomachischen Ethik) dargestellt hat, als Teil praktischer Vernunft die Fähigkeit, etwas Nützliches zweckgemäß herstellen zu können (›Handwerk‹), beruhend auf Fachwissen, Übung und Erfahrung, aufgrund von Einsicht in wirkliche Zusammenhänge (*technê* – ursprünglich eine List). Dieser Kerngedanke hat sich bis heute im Begriff der Technik erhalten, wie er etwa in Definitionen seitens der Technikphilosophie (Ropohl 1991), aber ebenso auch des VDI (VDI 1991) zum Ausdruck kommt. Entgegen einer heute oftmals sehr eingeschränkten Auffassung, die unter Technik lediglich eine Ansammlung von Artefakten versteht, wird darunter nicht nur die Gesamtheit von künstlichen Mitteln für gesellschaftliche Zwecke subsumiert. Vielmehr umfasst Technik, nach einem Bonmot von Ortega y Gasset verstanden als »Anstrengung, Anstrengungen zu ersparen«, neben zweckmäßig gestalteten Artefakten auch die kreativen Prozesse ihrer Konzeption und Herstellung wie auch die ihres Gebrauchs. Dazu gehören dann insbesondere auch Methoden und Verfahren der Herstellung sowie die Anstrengungen zur Aneignung ihrer Funktionen für den praktischen Gebrauch (vgl. »Grenzen und Widersprüche der Entwicklung und Anwendung ›Autonomer Systeme‹«).

Dabei gilt es, sich stets des fundamentalen Unterschieds zwischen den Fähigkeiten zur Konzeption und Herstellung der Artefakte und deren physisch verkörperten funktionalen Eigenschaften zum Erreichen der Zwecke als Ergebnis bewusst zu sein. Maschinelle bzw. technische Funktionen sind dem jeweiligen Zweck entsprechend genutzte Wirkungen der Natur oder, im Falle von Computern, durch binäre Signale in Schaltsystemen physisch verkörperte logische und arithmetische Operationen. Zu deren Konzeption bedarf es kreativer menschlicher Akteure – Konstrukteure und Entwickler –, die durch Einsicht in diese Zusammenhänge eine den jeweiligen Zwecken und Anforderungen genügende Konstellation von Funktionen der Artefakte in Form von Programmen zu schaffen imstande sind. Eben diese fundamentale Differenz zwischen tätigem Herstellen und sachlich Hergestelltem, zwischen Machen und Gemachtem, gerät aber in den Diskursen um »KI« oft aus dem Blick, wenn ständig Fähigkeiten zur Konzeption und Herstellung der Software als Artefakt mit Eigenschaften der Software selbst verwechselt werden. So müsste, um das Artefakt einer tatsächlich intelligenten oder von selbst lernenden Maschine konzipieren und deren Funktionsweise beschreiben zu können, zuvor analytisch verstanden worden sein, wie genau pure physiko-chemische neuronale Aktivität Bewusstsein und Reflexionsfähigkeit entstehen zu lassen und Kreativität hervorzubringen imstande ist.

Fünftens zeigt sich dabei oftmals auch ein unzutreffendes Verständnis von Maschinen. Als technische Artefakte vollziehen Maschinen mehr oder weniger selbsttätig zweckgemäß *gesteuerte*, kausal *determinierte wiederholbare Bewegungen*. Diese Bewegungen werden durch zuvor explizit erkannte natürliche Wirkungen bzw. logische Operationen erzwungen, die als zweckmäßig gestaltete und miteinander verknüpfte Funktionen physisch implementiert werden. Je nach Art und Implementierung dieser Bewegungen lassen sich zwei grundverschiedene Typen von Maschinen unterscheiden:

- Energie- und stoffumwandelnde Maschinen nutzen Natureffekte, indem sie durch thermo- oder elektrodynamische Wandlung potentieller Energie Kräfte erzeugen, um damit Antriebe zu realisieren (»Kraftmaschinen«) bzw. mittels mechanischer Funktionen Kräfte übertragen, um mithilfe der durch die Kräfte bewegten Werkzeugen Stoff zu formen (»Arbeitsmaschinen«); mithin ›maschinisieren‹ sie körperliche Arbeit.
- Semiotische (symbolische bzw. logische) Maschinen (»Computer«) operieren innerhalb von Zeichenprozessen sozialer Praxis: aufgrund zweckmäßiger formaler Modellierung bestimmter Vorgänge der Zeichenverarbeitung mittels bedeutungsloser Elemente eines Alphabets werden Algorithmen (berechenbare Funktionen) gewonnen, die logische und arithmetische Operationen programmgesteuert auszuführen erlauben; auf diese Weise ›maschinisieren‹ sie kognitive Arbeit.

Anstelle der thermo- bzw. elektrodynamischen in Verbindung mit mechanischen Wirkungen, die Bewegungen von Kraft- und Arbeitsmaschinen bestimmen,

steuern die Gesetze von Logik und Arithmetik die algorithmische Funktionsweise semiotischer Maschinen. Dementsprechend sind Computer Maschinen, deren mittels Modellierung und Formalisierung von Zeichenprozessen zweckmäßig konstruierte Algorithmen formale logische und arithmetische Operationen nutzen und ihrerseits in binären Schaltwerken als in ihrem Bewegungsablauf wohl definierte Abfolgen physischer Signalzustände ausgeführt werden.

In diesem Zusammenhang ist ferner der sehr bedeutsame Unterschied zwischen Modell und wirklicher Maschine wichtig. Ein Modell ist eine präzise und umfassende funktionale Beschreibung einer Maschine, während die Maschine selbst die physische Vergegenständlichung oder Realisierung des Modells ist. Ein Modell ist daher noch nicht die Maschine selbst, sondern nur deren Beschreibung: Bei physikalischen Maschinen hat die Beschreibung die Form von Zeichnungen und Stücklisten mit stofflichen Angaben, nach denen die Maschine erst noch gebaut werden muss; bei semiotischen Maschinen besteht im Unterschied dazu die funktionale Beschreibung aus formalen Datenstrukturen und Programmen, die aber – nach automatischer Übersetzung in Maschinencode – bereits maschinell ausführbar sind. Hier ist die formale Beschreibung, jedenfalls in lauffähiger Form, schon funktionaler Teil der Maschine.

Mithin ist die in der Computertechnik übliche Trennung von Soft- und Hardware letztlich irreführend und riskant, verführt sie doch ebenfalls zu illusionären Vorstellungen quasi ›kognitiver‹ Funktionen in Computern. Am Ende geht aber deren maschinelle Bewegung in getakteten Folgen von zweckmäßig gesteuerten physischen Signalzuständen auf: Das steuernde Programm wird dabei selbst als Binärcode gespeichert (wie auch die universelle Turingmaschine jede spezielle simulieren kann). Während der Laufzeit sind Computer als Maschinen reine Hardware; die Software dient lediglich der funktional eigenständigen Darstellung der Logik des Bewegungsablaufs und deren formale syntaktische Beschreibung als Algorithmus ermöglicht die maschinelle Ausführung. Die derart auf Computer-Schaltsystemen mittels elektronisch verkörperter kausaler Wirkungen realisierten logischen Funktionen entbehren somit jeglicher kognitiver Dimension: Computer ›wissen‹ nicht, was sie tun und wie sie es tun, noch ›wissen‹ sie, warum sie es tun (vgl. Winograd & Flores 1986).

Selbstredend sind viele der Aufsehen erregenden »KI«-Programme Ergebnis seriöser Forschung und Meisterleistungen kreativer Konstruktion von Berechnungsverfahren zur Lösung jeweils wohlbestimmter kognitiver Aufgaben. Deren Funktionsweise und Leistungsfähigkeit verdanken sich allein der programmgesteuerten Ausführung findig ausgedachter Algorithmen auf hochleistungsfähiger Hardware. Indem die so zustande gekommenen Ergebnisse jedoch von naiven Propagandisten ständig als neue Durchbrüche von »KI« mystifiziert werden, erhalten die Legenden »künstlicher Intelligenz« immer wieder neue Nahrung, während zugleich die eigentliche Leistung, die Raffinesse der Abstraktion, Formalisierung und Berechnung durch die Schöpfer der technischen Artefakte verborgen

bleiben. Selbstredend verkörpern Berechnungsverfahren Intelligenz, in ihnen ist die konkrete Intelligenz und Kompetenz ihrer Konstrukteure vergegenständlicht, die sie aber nicht selbst hervorbringen. Es ist wie beim Zaubern: Erst muss ein Kaninchen auf zu verbergende Weise in den Hut verbracht werden, aus dem es zum Erstaunen des Publikums wieder zum Vorschein kommt. Das gilt im übrigen für die Funktionsweise jedes technischen Artefakts. Niemand käme auf die verschrobene Idee, eine Waschmaschine oder einen Verbrennungsmotor »intelligent« zu nennen, obgleich beide über anspruchsvolle Steuerungen verfügen, die früher mechanisch, in neuerer Zeit auch algorithmisch durch Programme realisiert werden. Somit wird mit dieser Redeweise, die Computer-Artefakten menschliche Fähigkeiten zuschreibt, der grundlegende Unterschied zwischen dem absichtsvoll handelnden Urheber und dem Artefakt als Ergebnis seiner bewussten Tätigkeit verschleiert.

Aufgrund gänzlich unscharfer Definitionen kann bis heute niemand wissen, worin sich »KI«-Systeme überhaupt von anderen Computersystemen unterscheiden (wobei im allgemeinen Sprachgebrauch Computersysteme auch bereits als »eine KI« bezeichnet werden; vgl. den Beitrag ››Machines that think‹ – die »KI«-Illusion und ihre Wurzeln«). Solange diese Art Mystifikationen mangels analytischer Sorgfalt anstelle sachgerechter Darstellungen der tatsächlichen Funktionsweise fortgeschrittener Computersysteme stehen, lassen sich keine tragfähigen Erkenntnisse über den Einsatz dieser Artefakte und dessen Folgen gewinnen. Damit erfährt man weder etwas über deren Wirkungsweise noch etwas über deren Wirkungen im Gebrauch, noch lassen sich so gesellschaftliche Nebenwirkungen angemessen beurteilen. Mit dem ständigen Gebrauch vermenschlichender Metaphern für letztlich physische Vorgänge werden nur mystifizierende Legenden erzählt, aber kein zutreffendes Verständnis der technischen Artefakte gewonnen.

In Anbetracht dieser sich von Beginn an häufenden Missverständnisse erscheint es angebracht, grundlegende Einsichten der eigenen theoretischen Perspektive explizit offenzulegen. Bedauerlicherweise lässt genau das die Perspektive der »KI«-Adepten vermissen, sodass deren Prämissen erst wie oben skizziert aus bestimmten Äußerungen mühsam erschlossen werden muss. Es ist aber gute wissenschaftliche Praxis, grundlegende Annahmen und die darauf fussende Perspektive im einzelnen darzulegen.

3 Einsichten: Theoretische Perspektiven

3.1 Selbstorganisation: Natürliche und kulturelle Evolution

Natur ist in der hier vertretenen Perspektive alles, was sich ohne uns Menschen, ohne unser Wollen, durch Selbstorganisation von selbst vollzieht, ein Prozess natürlicher Evolution, der uns widerfährt. Uns Menschen, selbst ein Produkt der natürlichen Evolution, kennzeichnet dann die durch unsere bewusste Tätigkeit in der Natur gemeinschaftlich geschaffene Welt der *Kultur*. Mithin ist die kulturelle

Evolution die Fortsetzung der natürlichen mit anderen Mitteln. Deren Evolution vollzieht sich, angetrieben durch jeweils dominante Interessen, auf Basis kollektiv geteilter Intentionalität in Praktiken der Herstellung und des Gebrauchs von

- *Zeichen* als Mitteln sozialer Interaktion, Kommunikation und Reflexion (Sprache: vermittelt Bedeutungen),
- *Werkzeugen* als Mitteln zweckmäßiger Nutzung von Naturstoffen und -kräften (Technik: vermittelt Wirkungen).

Der Begriff des *Zeichens* und die Semiotik als Lehre vom Umgang mit Zeichen (Eco 1991) sind im vorliegenden Zusammenhang von grundlegender Bedeutung. Logisch am weitreichendsten analysiert und definiert wurde der Zeichenbegriff vom amerikanischen pragmatistischen Logiker C.S. Peirce (1983) als dreistellige Relation zwischen einem physischem Zeichenträger, dem damit sozial konstruiert bezeichneten Gegenstand oder Vorgang und deren Bedeutung in einer konkreten sozialen Praxis: *»Ein Zeichen ist etwas, das für jemanden in einer bestimmten Hinsicht oder Fähigkeit für etwas steht«* (vgl. unabhängig und äquivalent dazu Freges Begriffsschrift 1879).

Damit können in der Welt wahrgenommene Dinge oder Vorgänge bezeichnet und zugleich auf den Begriff gebracht, durch ein Urteil *über* sie *als etwas* erkannt (›prädiziert‹) werden – ein grundlegender Vorgang der Abstraktion. Durch die Bindung sozial konstruierter Bedeutung an einen physischen Zeichenträger vermag dieser Zeichenbegriff zudem zwischen Signal (in der physische Welt der Wirkungen) und Sinn (in der sozialen Welt der Bedeutungen) zu vermitteln und so eine enge Verbindung herzustellen zwischen Computertechnik und sozialer Praxis (vgl. den Beitrag »›Informatik‹ – eine Wissenschaft auf Abwegen« sowie »Falkenberg et al. 1998; Nake 2001).

Dabei ist zum Verständnis wichtig festzuhalten, dass sich im Verlauf der natürlichen Evolution zwei Entwicklungssprünge in der Selbstorganisation von Materie ergeben haben. Die Daseinsweise von Materie ist Bewegung und Selbstorganisation ist eine Bewegungsform fern vom Gleichgewicht, durch die aufgrund von Energiezufuhr unter bestimmten Umgebungsbedingungen dauerhaft höher organisierte Ordnungsstrukturen von selbst hervorgebracht werden. Beispielsweise erschaffen und erhalten sich lebende Organismen selbst (durch »Autopoiese«) und höhere Lebensformen entwickeln Bewusstsein:

- *Sprung ins Leben*: Übergang von der unbelebten zur belebten Natur (›Autopoiese‹),
- *Sprung ins Bewusstsein*: In der belebten Natur der Übergang von bloß sensiblen, aber unbewussten zu sozialen, sinngebenden und ihrer selbst bewussten Organismen (›Signifikation‹).

Während der erste Sprung inzwischen weitgehend aufgeklärt ist, liegt der zweite mangels einer Theorie des Bewusstseins noch größtenteils im Dunkeln. Wer aber die Unterschiede zwischen diesen Entwicklungsstufen ignoriert, wie das in »KI«-Diskursen oft geschieht, begeht mit der Verwechslung der Betrachtungsebenen

schwere Kategorienfehler und macht sich des Reduktionismus schuldig. So wird etwa oftmals das syntaktische ›Informationsmaß‹ (»Entropie«) der Shannonschen Signaltheorie in der physischen Welt mit dem Ergebnis von Signifikation durch Interpretation in sozialen Praktiken verwechselt (wovor dieser ausdrücklich gewarnt hat; vgl. Shannon 1948). Sinngebung, Zuweisung von Bedeutung, ist aber keine Leistung physikalischer, sondern sprachbegabter sozialer, mithin zu Intentionalität, Reflexion und Bewusstsein befähigter Systeme. Computer verarbeiten nur bedeutungslose Signale (logisch: ›Daten‹), keine ›Information‹.

In diesem Zusammenhang muss auch nochmals auf die aristotelische Unterscheidung der theoretischen (*epistêmê*) von der praktischen Vernunft (*technê*, *phronêsis* und *nous*) verwiesen werden. Die praktische Vernunft bzw. das lebendige Arbeitsvermögen als Inbegriff von Erfahrung und Können, von Urteilskraft und Analysefähigkeit ist natürlich und primär gegeben als Voraussetzung, um überhaupt theoretische Einsichten gewinnen, abstrahieren und Begriffe bilden sowie Probleme analysieren zu können. Deren Bedeutung für die Entwicklung von Wissenschaft und Technik haben gewichtige Stimmen gerade auch im Hinblick auf Computertechnik schon in frühen Phasen herausgestellt (vgl. etwa Dreyfus 1979; Ryle 1949/1987; Polanyi 1966/1985; Volpert 1999). Nur wird die Bedeutung dieses Arbeitsvermögens in »KI«-Diskursen immer wieder ignoriert.

Mit Hilfe von ›Zeichen‹ können Dinge oder Vorgänge der Lebenswelt bezeichnet (›benannt‹), *als etwas* begriffen, darüber hinaus in ihrem Zusammenhang gedacht und beschrieben und so insgesamt symbolisch nachvollzogen werden:

- zunächst zwecks Kooperation lautlich geäußert (zwecks gemeinschaftlicher Daseinsvorsorge durch Organisation des Stoffwechsels mit der Natur),
- später auch zwecks Verwaltung eines rasch wachsenden Mehrprodukts mittels Zahl- und Schriftzeichen auch schriftlich notiert (Flusskulturen um 3.000 v.u.Z.; Buchstaben stehen für Sprachlaute, Ziffern für gedachte (An-)Zahlen und Mittel zum Rechnen; *digitale Revolution*),
- zuletzt auch in Form binär codierter Datenstrukturen und Algorithmen (mit Methoden der Computing Science); Begreif- und Denkbares wird damit partiell berechenbar (*algorithmische Revolution*).

Der Zeichengebrauch der Sprache konstituiert die soziale Welt der Bedeutungen (im Unterschied zu den physischen Wirkungen der Natur) und erschließt durch Begriffsbildung Eigenschaften von Dingen oder Vorgängen (Taylor 2017). Im Unterschied zur physischen Welt werden Gegenstände und Tatsachen (›Fakten‹) der sozialen Welt – sog. *institutionelle Tatsachen* – erst durch Kommunikation und Kooperation als Formen zeichenbasiert koordinierten Handelns geschaffen und anerkannt. Sie beruhen auf kollektiver Intentionalität, d.h. auf geteilten Zielen und anerkannten komplementären Rollen, sog. Statusfunktionen, die durch konstitutive Regeln mittels Sprache – durch *deklarative Sprechakte* – geschaffen werden: *»Wir sorgen dafür, dass etwas der Fall ist, indem wir es als etwas repräsentieren, was der Fall ist«* (Searle 2012). Darin zeigt sich die ›performative Kraft‹

von Sprache (Austin 1962): Soziale Tatsachen werden buchstäblich ›ins Leben gerufen‹, sie sind menschengemacht, aber gesellschaftlicher Natur und damit überindividuell gegeben. Wichtige Beispiele sind etwa Spiele, die durch vereinbarte Regeln entstehen, das Geld, das kraft staatlicher Souveränität aufgrund gesetzlicher Zentralbank-Regeln funktioniert, oder die Mathematik, die mit nur gedachten Objekten operiert, die allein durch Alphabetzeichen und dafür geltende formale Regeln bestimmt sind – Hilbert zufolge ist sie *»ein Spiel mit wenigen Regeln und bedeutungslosen Zeichen auf Papier«*.

Wesentlich für das Verständnis von Prozessen sozialer Interaktion (Kommunikation und Kooperation) ist ferner, dass sich von selbst, gewissermaßen hinter dem Rücken der Akteure, bestimmte Regelmäßigkeiten oder Muster ausbilden, die sich analytisch in Vorgänge der Signifikation (Sinngebung), der Domination (Machtausübung) und der Legitimation (Sanktionierung) weiter unterscheiden lassen. Die Muster sind als Gewohnheiten Ergebnis und Voraussetzung künftigen Handelns zugleich: sie strukturieren das Handeln und ermöglichen Verständigung, determinieren beides aber nicht (die sprichwörtliche ›Macht der Gewohnheit‹, die sich auch brechen lässt). Soziale Praktiken sind regelmäßig, aber nicht kausal determiniert, mithin kontingent, sie verändern sich spontan und schleichend (vgl. *»Coping with Descartes' Error in Information Systems«*). Die Existenz der Muster und Gewohnheiten begründet einerseits, warum künftige soziale Praktiken in Grenzen vorhersehbar sind (eine wesentliche Erfolgsbedingung für den Einsatz konnektionistischer »KI«), während andererseits deren Kontingenz und schleichende Veränderung den Erfolg auf längere Sicht wieder untergräbt (Giddens 1988). Daher ist ein Vorgehen, das aus einem zum Zeitpunkt t quantitativ ermittelten Beziehungsgeflecht sozialer Praxis dessen Zustand zum Zeitpunkt $t + \Delta t$ zu berechnen sucht, von vornherein als fragwürdig zu betrachten. Dies umso mehr, als jeder praktische Gebrauch von Computertechnik selbstreferentiell auf die Praxis zurückwirkt, deren Modellierung er voraussetzt (vgl. mehr dazu im Beitrag *»Das Produktivitätsparadoxon der Computertechnik«*).

Eine Konsequenz der sozialen Konstruktion von Bedeutung ist, dass sich Zeichen nicht nur zur Abbildung realer Vorgänge, sondern auch zur Täuschung über sie verwenden lassen: *»Ein Zeichen ist alles, was sich als signifizierender Vertreter für etwas anderes auffassen läßt. … Also ist die Semiotik im Grunde die Disziplin, die alles untersucht, was man zum Lügen verwenden kann.«* (Eco 1991: 26). Aufgrund seiner Kreativität ist das menschliche Gehirn eben auch eine ›Fälscherwerkstatt‹. Insgesamt führt die Repräsentation von (ggf. auch irrtümlichem) Wissen über die physische und soziale Welt zu deren ›Verdoppelung‹ in einer (auch verzerrt) abbildenden Zeichen-Welt. Als sprachliche Modelle von Wirklichkeit bilden Beschreibungen eine eigene Welt, die sich freilich an der wirklichen Welt physischer und sozialer Gegebenheiten stets neu messen lassen muss: Wirklich sind nicht nur tatsächliche Ereignisse oder Vorgänge, sondern auch das, was mittels Zeichen über sie gesagt und gezeigt bzw. ›erzählt‹ wird, was

die Wirkmächtigkeit von ›Narrativen‹ ausmacht. Somit wirken Beobachtungen und Beschreibungen sozialer Praktiken bereits als Intervention. Sie benutzen dasselbe Medium der Sprache, mithin können sie verändern, was sie beschreiben und bilden damit eine Machtressource: Macht hat, wer bestimmt, was ›wirklich‹ ist, was den ›Raum des Sagbaren‹ abgrenzt – wobei freilich stets die Wirklichkeit selbst das ›letzte Wort‹ hat. Eingedenk dessen kommt der Aneignung und Realisierung methodisch gesicherter Erkenntnisse über Gegenstände und Vorgänge der physischen und sozialen Welt große Bedeutung zu, denn ohne sie droht ein gefährlicher Realitätsverlust. Dabei muss die Methoden der Ergebnissicherung den je besonderen Eigenschaften der Wissenschaftsbereiche Rechnung tragen – kontrollierte Experimente in den Naturwissenschaften, kontrollierte Praxistests in den Sozialwissenschaften und logische Beweisführung in der Mathematik (die ja nur mit gedachten Objekten operiert). Ohne so gesicherten Weltbezug ist man, wie u.a. »KI«-Diskurse belegen, den durch Erzählungen hervorgerufenen Illusionen schutzlos ausgeliefert.

Durch die Analyse realer Vorgänge und perspektivische Abstraktion lassen sich mittels Zeichen repräsentierte Denkvorgänge (›kognitive Funktionen‹) *partiell* (aber eben nicht vollständig) abbilden und formalisieren, in berechenbare Funktionen (Algorithmus) übersetzen und dann auch per Programm maschinell ausführen – auch Menschen rechnen, wenn sie einem formalen Verfahren folgen, wie Maschinen; ihr Denken ist aber nicht auf Berechnen beschränkt. Daher führt Computertechnik zur »Maschinisierung von Kopfarbeit« (Nake 1992), nicht zu »Künstlicher Intelligenz« (analog zur »Maschinisierung körperlicher Arbeit« mittels energie- und stoffwandelnder Maschinen).

Im Vergleich zu energie- und stoffwandelnden Maschinen offenbaren *semiotische* Maschinen erstaunliche Eigenschaften: Bei ihnen ist das ausführbare Programm deren (im Prinzip) lesbare Beschreibung wie auch deren operativer Kern zugleich. Umgekehrt können aus beschreibenden binär codierten Datenstrukturen wiederum algorithmisch (per Programm) Töne (Abtasttheorem), Bilder (Vektorgrafik) und Gegenstände (3D-Druck) physisch rekonstruiert werden, was mittels Schriftzeichen nicht möglich ist.

Der triadische Zeichenbegriff erweist sich somit für den praktischen Umgang mit Computertechnik als zentral. Er ist weit besser geeignet, die Funktionsweise von und Interaktion mit Computer-Artefakten zu analysieren und zu beschreiben als der in der Computertechnik leider häufig verwendete, hier aber irreführende Begriff der ›Information‹. Der Einsatz von Computertechnik in Zeichenprozessen sozialer Interaktion, zur Kommunikation und Kooperation, erfordert sowohl Beschreibungen in natürlicher Sprache als auch den Gebrauch formaler Sprachen (z.B. Programmiersprachen) zur Darstellung der Berechnungsverfahren. Dabei bestimmt die Grammatik einer formalen Sprache mit eindeutigen Regeln, welche der aus ihrem Alphabet gebildeten Wörter zulässig sind, während – anders als bei natürlicher Sprache – sinngebend interpretierende Aspekte, außer Betracht blei-

ben. Gleichwohl bezeichnet etwa die Academie Française die Lehre von der Computertechnik (Computing Science) mit dem Kunstwort ›Informatique‹ (ähnlich wie im Deutschen Informatik) als die *»Wissenschaft von der rationalen, insbesondere maschinellen Verarbeitung von Information«*. Computer operieren aber gerade nicht mit ›Information‹, sondern mit bedeutungslosen Signalen (physisch) bzw. Daten (logisch). So wird dem Begriff ›Information‹, dessen Konnotationen stets Vorgänge sozialer Deutung suggerieren, der maschinellen Verarbeitung bedeutungsloser Signale bzw. Daten erneut ein Bezug zu einer menschlichen Fähigkeit angedichtet (vgl. dazu mehr im Beitrag »›Informatik‹ – eine Wissenschaft auf Abwegen«).

3.2 Im Irrgarten der Selbstbezüglichkeit

Mit der Zeichenbildung hat sich der *homo sapiens* die generelle Möglichkeit geschaffen, beliebige Gegenstände, Vorgänge oder Sachverhalte zu bezeichnen und symbolisch zu repräsentieren, damit auch die Grundlage, das eigene Tun und Denken zu reflektieren, die Grundlage dafür, sich beides bewusst zu machen und in Gedanken probeweise zu handeln. Mit anderen Worten: das handelnde Subjekt hat sich die Voraussetzung geschaffen, sich selbst, sein Denken, Fühlen und Handeln, zu beobachten. Damit erschließt sich dem Organismus eine neue Dimension, die Möglichkeit, ein Bewusstsein über die Welt, über sich selbst und über sein Handeln in der Welt zu bilden.

So vermag das Subjekt die wichtige Unterscheidung zwischen sich selbst und der Welt als dem, was es nicht selbst ist, zu treffen – eine Unterscheidung, die, wie sich zeigt, hohe Ansprüche an logische Sorgfalt stellt. Das sich selbst reflektierende Subjekt muß dabei notwendigerweise zwei logische Betrachtungsebenen zugleich einnehmen: es ist sowohl denkendes (oder wissendes) Subjekt als auch gedachtes (oder gewusstes) Objekt. Als Subjekt vollzieht es die Handlung des Denkens an sich selbst; als Objekt ist es Gegenstand des Denkens, eine faktische Gegebenheit, die es sich selbst mitteilt. Es ist folglich subjektives, aktives Denken und objektives, passives Gedacht-Werden zugleich; förmlich ausgedrückt in den nachstehenden jeweils paarweise simultan geltenden Unterscheidungen:

Beobachten (Gegenstand) & Beobachten (Beobachten (Gegenstand)),
Denken (Gedachtes) & Denken (Denken (Gedachtes)).

Diese beiden Betrachtungsebenen miteinander zu einer Einheit vermitteln zu können, ist eine wesentliche Leistung des reflektierenden Subjekts, worin es seine Reflexionsidentität erfährt. Dieses Vermögen der Selbstreflexion unterscheidet den Menschen von weniger hoch organisierten Organismen, auf ihm beruhen u.a. auch sich seine Handlungskompetenz und seine Intelligenz, ferner seine Fähigkeit, aus Erfahrung zu lernen (Lernen lernen; vgl. z.B. Bateson 1985, Dreyfus & Dreyfus 1987; Nonaka 1994; Schon 1983).

Am Gebrauch der Sprache lässt sich diese Doppelbödigkeit bewussten Handelns demonstrieren. Sprache ermöglicht, Begriffe *über* Gegenstände oder

Sachverhalte der Welt *als* Zeichen zu vergegenständlichen (gesprochen oder geschrieben) und diese wiederum selbst zum Gegenstand von Begriffsbildung zu machen und somit Metazeichen, also Zeichen über Zeichen, zu bilden. In dieser Rekursivität des Zeichenbegriffs offenbart sich die Selbstbezüglichkeit der Zeichen: *»Jede natürliche Sprache dient permanent als Metasprache ihrer selbst durch jenen Prozess, den Peirce die unbegrenzte Semiose genannt hat«* (Eco 1994: 352). Sprache dient daher nicht nur als Medium der Kommunikation, sondern auch als Mittel der Reflexion und des Verstehens. Sie liefert insgesamt eine sprachliche Gliederung von Welt, womit wir die Welt als Wirklichkeit erfassen. Im sprachbegabten, damit zur Selbstreflexion befähigten Menschen hat sich die Natur die Möglichkeit geschaffen, sich partiell ihrer selbst bewusst zu werden.

Dabei muss in logischer Hinsicht sorgfältig unterschieden werden zwischen der *objektsprachlichen* Ebene mit Aussagen über nicht-sprachliche Gegenstände oder Vorgänge, die sich an den Wirkungen der Wirklichkeit messen lassen (müssen), und der *metasprachlichen* Ebene der grammatischen Regeln und prädikativen Aussagen *über* objektsprachliche Sätze. Wird diese Unterscheidung missachtet, kann man sich leicht in Probleme der Selbstreferenz verstricken, wie die nachstehende *Abb. 2* zeigt: Rechts ist zu sehen, wie Kommunikation mangels Unterscheidung zwischen Objekt- und Metasprache im gleichen Satz misslingt. Das Ergebnis entspricht übrigens genau dem eines Print-Befehls in einem Computerprogramm, in dem versäumt wurde, den Inhalt dem metasprachlichen Gebot entsprechend zuvor ins Arabische zu übersetzen.

(1) *»Dieser Satz ist falsch.«*

Diese metasprachliche Aussage wird zu einer Antinomie, falls sich ›dieser‹ auf den Satz selbst bezieht.

(2) *»Der Dorfbarbier rasiert alle Männer im Dorf, die sich nicht selbst rasieren.«*

Falls der Barbier im Dorf wohnt, entsteht die berühmte Russellsche Antinomie, die infolge naiver Mengenbildung die Mathematik 1903 in eine Grundlagenkrise stürzte:
Demnach gilt: $\forall x : (x \in \mathcal{R} \Leftrightarrow x \notin x)$.

Wird $\mathcal{R}$ für x eingesetzt, ergibt sich die Antinomie: $\mathcal{R} \in \mathcal{R} \Leftrightarrow \mathcal{R} \notin \mathcal{R}$.

Eine formal strenge Mengen-Axiomatik sorgt fortan dafür, dass nicht jedes Objekt Element einer Menge M sein kann, sondern M nur als $M = \{x \in y \mid \mathrm{A}(x)\}$ mit $\mathcal{R} \notin y$ existiert (Aussonderungsaxiom).

A new tanker arrives on site in Qatar:
The newly appointed American Manager
tells the local supervisor to ensure that
the tanker is clearly labeled:
"Diesel Fuel" in Arabic and
"No Smoking" in Arabic
This is what he got.

Abb. 2: Logische Probleme unbedachter Selbstreferenz (Foto: 9buz)

Auch bei formalen Sprachen und den bloß gedachten Gegenständen und Operationen der Mathematik muss die operative Betrachtungsebene des formalen Kalküls mit bedeutungslosen Alphabetzeichen sorgfältig von der Metaebene der Axiome und prädikativen Aussagen über formale Operationen des Kalküls unterschieden werden. Die *Abb. 2* zeigt links zwei prominente Beispiele dafür, wie es zu unauflösbaren Antinomien kommt, wenn Aussagen der Metaebene unbedacht auf Objekte der operativen Ebene bezogen werden und widersprüchliche Selbstreferenz entsteht. So hat die Russellsche Antinomie die Mathematik zu Beginn des 20. Jahrhunderts in eine Grundlagenkrise mit durchaus dramatischen Zügen gestürzt, die erst durch die sorgfältig axiomatische Begründung der Mengenlehre als Fundament der Mathematik überwunden wurde. Sie gab auch den Anstoß für das sog. »Hilbertprogramm«, das vorsah, die Mathematik auf der Basis plausibler, widerspruchsfreier Axiome als ein formales System zu definieren, innerhalb dessen die üblichen Beweismethoden gültig sein sollten (etwa nach dem Vorbild der Geometrie). Zu dessen Absicherung sollten außerhalb des Formalismus, auf der Ebene der Metamathematik, die Widerspruchsfreiheit der formal ableitbaren Sätze nachgewiesen werden. Zur allgemeinen Überraschung erwies sich das Programm jedoch in wesentlichen Teilen als nicht durchführbar.

Mithin widersetzt sich selbst die weitgehend formalisierte Mathematik ihrer lückenlosen Absicherung durch formale algorithmische Entscheidungsverfahren. So ist es erwiesenermaßen unmöglich,

- einen Algorithmus anzugeben, der alle Sätze eines formalen Systems der Stärke der Arithmetik abzuleiten und deren Widerspruchsfreiheit zu zeigen vermag (Gödel 1931);
- einen Algorithmus anzugeben, der von jeder Formel eines formalen Systems entscheiden kann, ob diese Formel ein wahrer Satz des Systems ist (Turing 1936).

Bezeichnenderweise beruht der Beweis von Gödel im Kern darauf, dass er als kompetenter Mathematiker metamathematische Prädikate *über* Terme des formalen Systems, hier die Beweisbarkeit, per Codierung als Formeln *im* System selbst auszudrücken vermag. Damit gelingt ihm, eine Formel im System so zu konstruieren, dass sie über sich als einem durch ihn als wahr erkannten Satz aussagt, nicht beweisbar zu sein. Zur mathematischen Fähigkeit gehört eben auch, dass sie bei allem, das sie operativ zu formalisieren vermag, durch Nachdenken über die Formalisierung und Intuition mittels abduktiven Schließens zu Einsichten gelangen kann, die außer Reichweite der Formalisierung liegen. Eben dieser kreative Akt ist ein Ausweis natürlicher Intelligenz.

Dagegen gibt die Abfolge von Operationen eines Algorithmus zwar Auskunft auf die Frage, was dabei im einzelnen abläuft; sie beantwortet aber nicht die Frage nach deren Sinn oder Bedeutung, warum sie so ablaufen. Operationen sagen nichts über sich selbst aus, etwa ob sie korrekt oder gebrauchstauglich sind. So ist etwa die Frage, ob ein Programm nach endlich vielen Schritten terminiert (oder

auch andere nicht-triviale Eigenschaften hat), formal per Algorithmus nicht entscheidbar.

Die nachstehende Erläuterung skizziert die Kerngedanken zum Gödelschen Beweis seines Unvollständigkeitssatzes und demonstriert eindrücklich das Zusammenspiel der Ebene des operativen Kalküls mit der Ebene metamathematischer Aussagen über sie (vgl. Hofstadter 1979; genauere Einzelheiten finden sich bei Hoffmann 2013):

Gödels tragende Idee zum Beweis der Unvollständigkeit der Arithmetik ist, zunächst einen arithmetischen Kalkül zu entwickeln, mittels dessen sämtliche Zeichen, Axiome, Sätze und Beweise, mithin alle Formeln des Systems umkehrbar eindeutig durch natürliche Zahlen ausgedrückt werden können. Damit ist es möglich, auch metamathematische Aussagen über Sätze des Systems in diesem System selbst abzubilden. Auf diesem Wege konstruiert Gödel einen Beweisgang, der im System unentscheidbare Sätze erzeugt. Mit Hilfe der Abbildung metamathematischer Sätze in arithmetische Formeln konstruiert er eine selbstbezügliche arithmetische Formel in Gestalt einer Kette von Alphabetzeichen, die sich so interpretieren läßt, dass sie über sich selbst etwas aussagt.

Das formalisierte System verwendet als Alphabet ausschließlich folgende Zeichen:

Logische Zeichen: ¬ [nicht], & [und], [,], (*x*) [„für alle *x:* ...“]

Arithmetische Zeichen: = [gleich], + [plus], · [mal], *x* [Variable],

'[Nachfolger von]

Zahlzeichen: ('), (''), ('''), ... [1, 2, 3, …]; *x, x', x''*, ... [verschiedene Variable].

Mit diesen Zeichen können sämtliche Formeln des Systems erzeugt werden. Jedem dieser Zeichen wird nun umkehrbar eindeutig eine der neun Ziffern wie folgt zugeordnet:

¬	&	(	)	=	+	·	*x*	'
1	2	3	4	5	6	7	8	9

Damit kann jede Formel des Systems durch die Ziffernfolge einer natürlichen Zahl – die sog. Gödelnummer – umkehrbar eindeutig codiert werden: zu jeder Formel gehört genau eine Zahl, ihre Gödelnummer, und zu jeder Gödelnummer gehört genau eine Formel. Z.B. wird der Satz »Für alle x, x' gilt: $x' + x = x + x'$« durch die Formel $(x)\ (x')\ (x' + x = x + x')$ ausgedrückt, zu der die Gödelnummer 384389438968586894 gehört. Mit diesem »Gödelisierung« genannten Verfahren gelingt es, infolge der umkehrbar eindeutigen Zuordnung einer natürlichen Zahl zu einer Formel des Systems, diese als arithmetischen Ausdruck darzustellen und umgekehrt einen arithmetischen Ausdruck als Formel oder Satz zu interpretieren.

Insbesondere werden damit auch metamathematische Aussagen über Formeln des Systems als arithmetische Beziehungen zwischen Gödelnummern ausgedrückt. Beispielsweise wird die metamathematische Aussage »Die Formelfolge mit der Gödelnummer *x'* ist ein Beweis für die Formel mit der Gödelnummer *x*« dargestellt als die arithmetische Beziehung B (*x', x*), ein sog. Beweispaar. Wichtig

ist ferner noch die Substitutionsbeziehung GS (x', x), die jede freie Variable einer Formel durch die Gödelnummer dieser Formel ersetzt; sie liefert die Gödelnummer, die durch diese Substitution erzeugt wird; z..B. hat die Formel $x = x$ die Gödelnummer 858; durch Ersetzen der Variablen x durch diese Nummer erhält man die Gödelnummer 8585858.

Mit diesen Hilfsmitteln kann nun die nachstehende Formel G' gebildet werden:

$$\neg B\ (x', x)\ \&\ GS\ (x'', x). \qquad (G')$$

Sie hat eine Gödelnummer n (deren genaue Ziffernfolge hier nicht weiter interessiert, geschrieben als n-maliges '-Zeichen: ' ' ... '), mittels derer die Substitutionsbeziehung GS für diese Formel mit der einzigen freien Variablen x'' durchgeführt werden kann und so die neue Formel G liefert:

$$\neg B\ (x', x)\ \&\ GS\ (' \ ' \ ... \ ', x). \qquad (G)$$

Diese Formel G besagt: Die Formel, deren Gödelnummer durch die Substitutionsbeziehung GS erzeugt ist, ist nicht beweisbar. Laut Definition ist diese durch GS erzeugte Gödelnummer diejenige der Formel G; daher lautet die Interpretation von G auch: Die Formel G ist nicht beweisbar, sie behauptet mittels der Gödelisierung ihre eigene Unbeweisbarkeit. G ist aber doch eine wahre Aussage, indem sie die Nichtbeweisbarkeit behauptet. Das formale axiomatische System der Arithmetik ist folglich unvollständig.

Dieser Beweisgang nutzt die mittels der ›List‹ der Gödelisierung geschaffene Möglichkeit, einen Satz der Metamathematik *über* das System als eine Formel *im* System auszudrücken, damit Selbstreferenz derart herzustellen, dass eine arithmetische Formel ihre eigene Unbeweisbarkeit aussagt. Die intuitive Konstruktion dieser Formel verweist auf spezifisch menschliche Kreativität und Intelligenz.

Es folgt dann noch der Beweis für die Unbeweisbarkeit der Widerspruchsfreiheit mit dem System eigenen Mitteln (der hier nicht weiter interessiert).

Im Lichte der Unvollständigkeit formaler Systeme hinreichender Mächtigkeit erweist sich die Computertechnik, zumindest was die grundlegenden theoretischen Einsichten der Computing Science über Formalisierung, Aufzählbarkeit, Entscheidbarkeit und Berechenbarkeit anbelangt, als eine Frucht der Grundlagenkrise der Mathematik und ihrer Überwindung mit freilich überraschendem Ausgang. Jedenfalls gehört sie mit ihren logisch-arithmetischen Operationen und deren physischer Implementierung als ›semiotische‹ Maschine der Objektwelt nichtsprachlicher Gegenstände und Vorgänge an. Dagegen bilden die in Form objektsprachlicher Aussagen gefassten Erzählungen über »KI«, wie oben gezeigt, allenfalls rein sprachlich deklarierte institutionelle Tatsachen, allerdings ohne jeden Bezug auf die tatsächliche Funktionsweise der als solchen nur bezeichneten technischen Artefakte. Damit sind sie als sprachlich generierte Konstrukte eine Illusion – Gegenstand von Selbsttäuschung, die sich zu dauerhafter Verblendung verfestigt: *»Der meiste Schaden, den der Computer potenziell zur Folge haben könnte, hängt weniger davon ab, was der Computer tatsächlich kann oder nicht kann,*

als vielmehr von den Eigenschaften, die das Publikum dem Computer zuschreibt. Der Nichtfachmann hat überhaupt keine andere Wahl, als dem Computer die Eigenschaften zuzuordnen, die durch die von der Presse verstärkte Propaganda der Computergemeinschaft zu ihm dringen.« (so Weizenbaum schon 1972 anlässlich der Einrichtung des Informatik-Fachbereichs an der Universität Hamburg, vgl. Die ZEIT 03/1972: 43).

Vor diesem Hintergrund mutet die Vorstellung, technisch von Ingenieuren konstruierten Berechnungsverfahren »Intelligenz« zuzuschreiben, doch eher skurril an. Ironisch zugespitzt erscheint die unterliegende Überzeugung, geordneten Abfolgen abstrakter logisch-arithmetischer Operationen wohne eine Art Geist inne, der dem Prozess der Ausführung dieser Operationen zu eigener »Intelligenz« verhelfe, als eine Art postmoderner Animismus. Die historischen Animisten glaubten fest daran, die Dinge der Natur seien beseelt und sie sahen sich daher gezwungen, die diesen innewohnenden bösen Geister zu besänftigen und die guten zu ihren Gunsten wohl gesonnen zu stimmen. Ganz ähnlich sind heute postmodern ›gebildete‹ »KI«-Adepten davon überzeugt, der komplizierten programmierten Abfolge formaler logischer und arithmetischer Operationen und ihrer Ausführung auf binären Schaltsystemen wohne eigene Einsichts- und Lernfähigkeit inne, eben eine Art »Intelligenz«. Dieser Glaube steigert sich bei den Anhängern des Transhumanismus (Bostrom 2016) sogar zu dem Wahn von einer technisch realisierbaren ›Unsterblichkeit‹, indem sich der eigene Geist dauerhaft auf Computer-Hardware ›scannen‹ lasse. Zudem werde schon in Kürze mit der weiter rasch wachsenden Rechenleistung und schier unbegrenzter rekursiver Berechnung der Punkt der »Singularität« erreicht, ab dem die »künstliche Intelligenz« maschinellen Rechnens die lebendige Intelligenz gewöhnlicher Menschen übertreffe (Kurzweil 2006). Diese reduktionistische und nekrophile Technikdystopie ist ohne Zweifel der Gipfelpunkt postmoderner Verblendung – eine Art Regression in prä-Galileische Zeiten, in der bekanntlich Wissenschaft und Magie nicht getrennt, sondern vielfältig verbunden waren und Konstrukteure von Automaten als Vertreter ›schwarzer Magie‹ galten.

So offenbart sich im öffentlichen Diskurs um »KI« (wie im übrigen auch in vielen anderen) letztlich nur ein weiteres Feld diskursiv erzeugter Verblendung. Mangels Bezug auf die wirklichen, nicht-sprachlichen logisch-arithmetischen und physisch-maschinellen Operationen verlieren sich die irreführend von anthropomorphisierenden Metaphern geprägten Aussagen im Dickicht postmoderner Narrative (im ›Rhizom‹). Im ständigen Kampf zwischen Vernunft und Verblendung scheint, wie die Geschichte lehrt, die aristotelische praktische Vernunft oftmals den kürzeren zu ziehen und Verblendung zunächst den Sieg davonzutragen: Zwar hat die Evolution den *homo sapiens* mit der Gabe reflexiver Vernunft ausgestattet, was aber nicht bedeutet, dass er gezwungen ist, davon auch Gebrauch zu machen. Die Art, wie große Teile der Gesellschaft vermeintliche »KI-Erfolge«, etwa AlphaGo oder GPT, wahrnehmen, sagt zwar nichts über die den Systemen zuges-

chriebene »Intelligenz«, aber umso mehr über die Unvernunft vieler Diskursteilnehmer aus, die lediglich eigene Fähigkeiten auf Computer-Artefakte projizieren. Vermeintliche Systemerfolge erweisen sich zudem immer wieder als Pyrrhussieg, denn auf längere Sicht triumphiert die Dominanz der Wirklichkeit mit ihren unwandelbaren logischen Operationen und physischen Wirkungen, die sich zwar partiell erkennen und ›listig‹ für eigene Zwecke nutzen, aber nicht zur Gänze ignorieren lassen.

Obgleich sich seit den frühen physischen Realisierungen programmierbarer rechnender Maschinen keine grundlegend neuen theoretischen Einsichten ergeben haben – abgesehen von Fortschritten in der Softwaretechnik und enormen Steigerungen der Rechenleistung der Hardware, die umfangreichere Berechnungen ermöglichen – sind jüngst die Diskurse um »KI« wieder mächtig angeschwollen. Damit wird erneut einem technologischen Determinismus gehuldigt (der seit den 1980er Jahren als überwunden galt). Mit deren gesellschaftlicher Dominanz wird freilich der Blick auf alternative Entwicklungsperspektiven der Computertechnik verstellt, die sich auf Dauer jedoch als deutlich ertragreicher erweisen könnten.

Idealtypisch stehen von Beginn an zwei grundverschiedene Entwicklungsperspektiven im Raum – im Englischen in Gestalt eines Wortspiels eindrücklich gekennzeichnet durch die Abkürzungen *AI* (für *»artificial intelligence«*) versus *IA* (für *»intelligence augmentation«*):

- Einerseits wird unter der *technikzentrierten »AI«*-Perspektive der Entwicklung »künstlicher Intelligenz« mit enormen Forschungsanstrengungen zur Realisierung sog. »autonomer Agenten« versucht, das einzigartige menschliche Arbeitsvermögen vermeintlich soweit nachzuahmen, um es in kognitiven Tätigkeiten weitgehend ersetzen zu können.
- Andererseits richten sich unter der *praxistheoretischen »IA«*-Perspektive die Forschungsanstrengungen darauf, lebendiges Arbeitsvermögen, seine Kompetenz, Intelligenz und Kreativität mit maschineller Rechenleistung produktiv zu verbinden und durch den Einsatz von aufgabengemäß ausgeklügelten Berechnungsverfahren so zu unterstützen, dass in ihrem Zusammenwirken höhere Produktivität und Innovationsfähigkeit bei der Aufgabenbewältigung entsteht und sich das Arbeitsvermögen dabei zugleich weiter zu entfalten vermag (infolge der Dynamik der Explikation von Praxis in Wissen, Konstruktion von Artefakt-Funktionen und deren Aneignung zu leistungsfähigerer Praxis; vgl. dazu Kap. 4 im Beitrag »Industrie 4.0 und Big Data – wirklich ein neuer Technologieschub?«).

Erstaunlicherweise und entgegen den Hypes um »KI« hat sich die Entwicklung der Computertechnik der letzten inzwischen acht Dekaden tatsächlich ganz überwiegend an der zweiten, der »IA«-Perspektive, orientiert und in verschiedenen Anwendungsbereichen große Erfolge zu verzeichnen. Große Teile kognitiver kooperativer Arbeit lassen sich relativ leicht durch formale Modelle beschreiben und mittels Berechnungsverfahren organisatorisch restrukturieren. Naheliegende Bei-

spiele dafür sind zunächst einmal alle jene Felder konstruktiver oder steuernder Ingenieurarbeit, die immer schon weitreichend durch Berechnungen unterstützt wurden (Dynamik von Bewegungen, Festigkeit und Stabilität von Strukturen, Planungsaufgaben (Operations Research), aber etwa auch logistische Prozesse und Vorgänge der Dokumentverwaltung. In kulturhistorischer Perspektive waren es gerade die Herausforderungen der Verwaltung eines rasch wachsenden gesellschaftlichen Mehrprodukts in den städtischen Flusskulturen im 3. Jahrtausend v.u.Z., die neben astronomischen Beobachtungen die frühe Entwicklung von Berechnungsverfahren vorantrieben. Ebenso sind die rasanten wissenschaftlich-technischen Fortschritte der Neuzeit engstens mit der Entwicklung von Berechnungsverfahren der Analysis und linearen Algebra verknüpft, deren Gebrauch wiederum diesbezügliche Arbeitsprozesse tiefgreifend verändert hat (jüngstes Beispiel: FEM der theoretischen Mechanik). So gehören etwa Dokumenteditoren oder CAD als Werkzeuge, ERM- und EDM-Systeme (Enterprise Resource und Document Management) als Organisationsmedien, das Internet mit WWW und CSCW als Medien der Kooperation zu den am weitreichendsten und verbreitetsten genutzten Entwicklungen (einschließlich E-Mail als weltumspannendes, vollständig automatisiertes Postsystem).

Computer science is the study of what can be computed and how to compute it.

Computational thinking is a fundamental skill for everyone, not just for computer scientists. It involves **analytical efforts for solving problems** and **designing artificial systems** for interacting with the physical or social world:

- ***Conceptualizing, not programming***: Thinking on different levels of abstraction for being able to program a computer.
- ***Understanding that humans, not computers, think***: Looking for ways how humans analytically solve problems rather than trying to get humans to think like computers performing a mechanical routine.
- ***Combining mathematical with engineering thinking***: It draws on mathematical thinking as its formal foundation; it draws on engineering thinking to build systems to interact with the real world.
- ***Dealing with ideas, not artifacts***: It consists of computational concepts humans use to approach and solve problems, to communicate and interact with other people.
- ***Developing a common skill for everyone everywhere***: This ability becomes a reality when it is so integral to human endeavors that it disappears as an explicit philosophy.

Abb. 3: ›Computational Thinking‹ als grundlegende Fähigkeit (Wing 2006)

Im Unterschied dazu rufen spektakulär erscheinende »KI«-Erfolge wie die Weltmeisterschaft im GO-Spiel durch AlphaGo Zero oder ChatGPT von OpenAI zwar große Verblüffung hervor, aber eher keinen oder nur geringen gesellschaftlichen Nutzen. Im Einzelfall, bei sehr speziellen Aufgaben, z.B. der Klassifikation von Prüflingen in der Qualitätssicherung, können KNN durchaus von Nutzen sein; das ist aber im Einzelfall methodisch sorgsam zu prüfen und ist aus dargelegten

Gründen nicht generell zu erwarten. Um das beurteilen zu können, bedarf es freilich einer grundlegenden, aber derzeit nur gering entwickelten Fähigkeit berechnungsbezogenen Denkens auf Seiten aller Beteiligter wie sie in *Abb. 3* näher gekennzeichnet ist. Ohne sie bleiben weiteren Mystifikation Tor und Tür geöffnet.

So steht am Ende der gesellschaftliche Umgang mit Computertechnik am Scheideweg: Statt sich ständig durch irregeleitete Vorstellungen über Leistungen von Berechnungsverfahren verführen und verblüffen zu lassen, das weder deren Zustandekommen sachgerecht zu beschreiben, noch über entsprechende menschliche Fähigkeiten Auskunft zu geben vermag, wäre es sehr viel fruchtbarer, danach zu fragen, wie sich tatsächlich drängende gesellschaftliche Probleme mittels welcher zu konstruierender Berechnungsverfahren auf Basis verfügbarer Rechenkapazität erfolgversprechend bewältigen lassen. Realistische Einschätzungen von Aufwand und Nutzen müssen dabei wie stets bei technischen Entwicklungen den Weg weisen.

4 Beiträge zur Kritik konnektionistischer »KI«

Die dargelegten Prämissen und Perspektiven bilden den Hintergrund der einzelnen Beiträge zu diesem Buch und stellen auch Zusammenhänge zwischen ihnen her. Sie sind im folgenden nicht chronologisch nach dem Zeitpunkt ihres Erscheinens angeordnet, sondern nach inhaltlichen Gesichtspunkten. Die folgenden kurzen Übersichten geben Auskunft über Anlass und Kontext ihrer Entstehung sowie über wesentliche Aspekte ihres Inhalts.

(1) Am Anfang steht zur Vertiefung grundlegender theoretischer Perspektiven der Beitrag *»Coping with Descartes' Error in Information Systems«*. Er entstand 2018 aus Anlass des Todes von Hubert Dreyfus für »AI & Society. Journal of Knowledge, Culture and Communication«, das aus diesem Anlass ein Themenheft publiziert hat. Darin werden mit der praxis- bzw. »kontakttheoretischen« Perspektive des leiblichen »being-in-the-world« und der repräsentationalen, auf begriffliche Erkenntnis fokussierenden Perspektive zwei grundverschiedene Zugänge des Menschen zur äußeren Welt dargestellt, die für lange Zeit abendländische Denktraditionen geprägt haben.

Beide Perspektiven haben gewisse eigene Berechtigung, im Fokus steht hier aber die Dynamik der Wechselwirkung ihrer jeweiligen Produkte. So wird zum besseren Verständnis von Menschen und Maschinen insbesondere analysiert, wie beide Weltzugänge zueinander in Beziehung stehen und im Ergebnis zusammenwirken, genauer: wie explizites begrifflich-propositionales Wissen partiell aus Erfahrungswissen und implizitem Können gewonnen, zur Konstruktion von Artefakten genutzt werden und diese umgekehrt wiederum durch Aneignung für den praktischen Gebrauch Erfahrung und Handlungskompetenz erweitern können.

Diese Dynamik der Explikation von Können und Erfahrung in begriffliches Wissen und dessen Aneignung für erweiterte Handlungskompetenz wird zur Ana-

lyse von Entwicklung und Gebrauch von Computerartefakten als semiotischen Maschinen genutzt. Dabei kommen dem Zusammenspiel von anschaulichem und begrifflichem Denken und der oftmals vernachlässigten Schlussweise der »Abduktion« besondere Bedeutung zu, wie sie vom amerikanischen pragmatistischen Logiker Peirce (1934) untersucht worden ist. Es ist ein logischer Schluss, der aufgrund menschlicher Intuition und Kreativität die Bildung von Hypothesen zwecks Erklärung unverstandener Phänomene erlaubt, aber noch methodisch gesicherter Prüfung bedarf; mit ihr werden die bekannten Schlüsse der Deduktion relevanter Prüfbedingungen und der Induktion zur Interpretation von Prüfergebnissen zu einem logischen Dreischritt der Erkenntnisgewinnung erweitert.

(2) Der zweite Beitrag mit dem Titel: *»›Super-intelligent‹ Machine: Technological Exuberance or the Road to Subjection«,* bereits 2017 im selben Journal erschienen, setzt sich anlässlich des deutlichen Anschwellens der dritten »KI«-Welle konkret mit Anspruch und Wirklichkeit der zugrunde liegenden ›konnektionistischen‹ Ansätze auch auf funktionaler Ebene kritisch auseinander. Dabei liegt ein besonderes Augenmerk auf dem Einsatz von Multi-Agentensystemen in Produktionsprozessen.

Bei diesen Ansätzen stehen und fallen die Güte und damit auch der Nutzen der Ergebnisse des Einsatzes dieser Systeme mit der Qualität der Daten, die zum ›Training‹ der Netzwerkstruktur, d.h. zur Berechnung der Netzwerk-Gewichte, verwendet werden. Herkunft und Qualität dieser Daten sind aber oftmals hinsichtlich Korrektheit, Aktualität und Repräsentativität fragwürdig oder schwer einzuschätzen. Zudem können mit diesen konnektionistischen Ansätzen stets nur wahrscheinliche und folglich keine sicheren Ergebnisse erzielt werden, wobei zudem methodisch fraglich ist, inwieweit bei kontingenten, sich verändernden sozialen Praktiken überhaupt mit Daten aus der Vergangenheit zukünftiges Verhalten hinreichend bestimmt werden kann. Diese Unsicherheiten beim praktischen Einsatz der Systeme bereiten den Nutzern beträchtliche Probleme und Belastungen (die in einem weiteren Beitrag noch gesondert untersucht werden, vgl. (6)).

Eingedenk dieser nur schwer behebbaren Schwierigkeiten plädiert der Beitrag dafür, mit größerer Aussicht auf Erfolg der alternativen praxistheoretischen Entwicklungsperspektive der »Intelligenzverstärkung« zu folgen und lebendige Arbeit durch Berechnungsverfahren zu unterstützen.

(3) Der an dritter Stelle folgende Aufsatz *»Industrie 4.0 und Big Data – wirklich ein neuer Technologieschub?«* ist ein aktualisierter Beitrag zur überarbeiteten zweiten Auflage des 2018 von Hartmut Hirsch-Kreinsen, Peter Ittermann und Jonathan Niehaus herausgegebenen Buches zur *»Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen«*. Ausgehend von damaligen Diskurs-Schwerpunkten und der rasch wachsenden Bedeutung und Entwicklungsdynamik kooperativer Wissensarbeit in der Produktion wird untersucht, worin sich die neueren Ansätze der modisch so genannten »In-

dustrie 4.0« von früheren Bemühungen um eine flexible Automatisierung von Produktionsprozessen mittels Computerartefakten unterscheiden.

Dabei wird gezeigt, dass sich die an die künftige technisch-organisatorische Entwicklung von Produktionsprozessen marktseitig zu stellenden Anforderungen hoher Flexibilität bei zugleich hoher Produktivität gegenüber früher in keiner Weise geändert haben, wohl aber die bevorzugten Lösungsansätze: Konzentrierten sich aufgrund der Erfahrung des Scheiterns der auf symbolischer »KI« beruhenden technikzentrierten Ansätze des »Computer-integrated Manufacturing (CIM) voraufgehende Entwicklungsanstrengungen vorwiegend auf die »holonische« Reorganisation der Prozesse mit Unterstützung durch Computerartefakte, so werden mit »Industrie 4.0« erneut überwiegend technikzentrierte Konzepte verfolgt, die der Beitrag kritisch unter die Lupe nimmt.

In diesen Konzepten spielen abermals die schon behandelten »Multi-Agentensysteme« auf Basis von KNN eine zentrale Rolle, die jeweils für ihre besondere Aufgabe mittels großer Datenmengen (»Big Data«) ›trainiert‹ werden müssen. Folglich gelten auch hier die schon angesprochenen methodisch-technischen Grenzen solcher Systeme aufgrund fehlender oder unsicherer Qualität der Daten und der aus der Unsicherheit der Ergebnisse resultierenden Probleme in ethischer Hinsicht, aber auch im Hinblick auf Behinderungen der Entfaltung von Arbeitsvermögen.

(4) Im vierten Aufsatz mit dem Titel: *»Grenzen und Widersprüche der Entwicklung und Anwendung ›Autonomer Systeme‹«*, der in dem 2019 von Hartmut-Hirsch-Kreinsen und Anemari Karačić herausgegebenen Buchs über *»Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt«* erschienen ist, werden die Funktionsweise von KNN als dominanter technischer Basis der neuen »KI« und des sog. »Deep Learning« sowie darin begründete inhärente Probleme auf technisch-funktionaler Ebene vertiefend betrachtet und anhand von Beispielen aufgezeigt. Dabei hängen die Erfolgsaussichten des Einsatzes dieser Systeme wesentlich von der Art der Aufgabe ab, zu deren Lösung sie eingesetzt werden. »KI«-Systeme sind keineswegs das Allheilmittel, für das sie häufig gehalten werden.

Vor dem historischen Hintergrund der Entwicklung der ideellen Grundlagen der Computertechnik, deren sträfliche Vernachlässigung in »KI«-Diskursen Ursache vieler Missverständnisse dieser Ansätze und der Fehleinschätzung ihrer Leistungsfähigkeit ist, wird zudem gezeigt, dass die in diesem Zusammenhang verwendete Sprache durchweg auf unsachgemäßen und irreführenden Metaphern beruht. Dadurch werden menschliche Fähigkeiten einfach auf programmierte, maschinell ausgeführte Abläufe projiziert, ohne dass dafür weder Anlass noch Berechtigung in der Sache gegeben wären.

(5) Diese Perspektive wird mit dem Beitrag: *»›Machines that think‹ – die »KI«-Illusion und ihre Wurzeln«* zu dem von Klaus Lenk und Jörg Pohle 2021 herausgegebenen Buch mit dem Titel: *»Der Weg in die ›Digitalisierung‹ der*

Gesellschaft. Was können wir aus der Geschichte der Informatik lernen?« weiter vertiefend verfolgt.

Dabei müssen schon die unterschiedlichen Versuche, »KI«-Systeme zu definieren und von anderen Computersystemen zu unterscheiden als kläglich gescheitert angesehen werden. Infolgedessen kann eigentlich niemand wissen, wovon bei »KI« tatsächlich die Rede ist. Der vorherrschenden Mystifizierung der Computertechnik und insbesondere ihrer ›Vermenschlichung‹ mittels durchweg unpassender Metaphern lässt sich nur entkommen, wenn im einzelnen analysiert wird, wie deren technische Funktionalität zustande kommt und wo prinzipielle Grenzen der Formalisierung und Modellierung verlaufen. Die vermeintliche ›List‹, mit Methoden des sog. »Deep Learning« die Mühen der angemessenen Modellierung umgehen zu können, werfen geradenwegs noch gravierendere Probleme auf (wie etwa das der in (6) abgehandelten psychischen Belastungen).

(6) Ein Kernproblem im praktischen Umgang mit konnektionistischen »KI«-Systemen sind die sehr hohen psychischen Belastungen, die von der »Ko-Aktion« mit den sich prinzipiell unsicher verhaltenden Systemen hervorgerufen werden. Dieses Kernproblem wird im 6. Aufsatz: *»Paradoxien der Ko-Aktion von Experten und adaptiven Systemen«* untersucht, der in dem 2020 vom Autor gemeinsam mit Klaus Fuchs-Kittowski herausgegebenen Tagungsband zur Konferenz der Leibniz-Sozietät zum Thema: *»Zukunft der Arbeit – soziotechnische Gestaltung der Arbeitswelt im Zeichen von ›Digitalisierung‹ und ›Künstlicher Intelligenz‹«* erschienen ist.

Darin werden – vor dem Hintergrund der bereits behandelten grundsätzlichen Herausforderungen im Umgang mit adaptiven Systemen wie KNN – mittels eines empirisch seit langem erprobten relationalen Modells der Genese psychischer Belastungen die besonderen Belastungen untersucht, die unter Leistungsdruck aus der »Ko-Aktion« mit diesen Systemen und ihren unzuverlässigen Ergebnissen resultieren. Darüber hinaus werden daraus Anforderungen an den Einsatz der Systeme abgeleitet, die bis zu einem Verbot reichen, solange sie noch keine hinreichenden Erklärungen für ihr Verhalten in einer spezifischen Arbeitssituation geben können.

(7) Der Einsatz von »KI«-Systemen wird häufig – wie übrigens auch schon bei ›gewöhnlichen‹ Computersystemen zuvor – mit weitreichenden Erwartungen großer Produktivitätssteigerungen, nun auch im Bereich der Wissensarbeit, und dementsprechend mit Befürchtungen um Beschäftigungsverluste oft apokalyptischen Ausmaßes in Verbindung gebracht. Diesen Zusammenhängen geht der Aufsatz mit dem Titel: *»Das Produktivitätsparadoxon der Computertechnik«* nach, der in dem 2020 von Heinz J. Bontrup und Jürgen Daub herausgegebenen Buch zum Thema: *»Digitalisierung und Technik – Fortschritt oder Fluch? Perspektiven der Produktivkraftentwicklung im modernen Kapitalismus«* erschienen ist.

Darin wird zunächst empirisches Material zum Produktivitätsparadoxon präsentiert, demzufolge in entwickelten Volkswirtschaften auf gesamtwirtschaftlicher Ebene durch Computereinsatz keine zusätzlichen Produktivitätssteigerungen ausgelöst werden und auf einzelwirtschaftlicher Ebene nur dann, wenn sie mit tief greifenden organisatorischen Restrukturierungen und massiver Qualifizierung einhergehen. Anschließend wird auf verschiedenen Betrachtungsebenen nach Erklärungen für dieses empirisch gut gesicherte Phänomen gesucht.

Das beginnt schon auf der technischen Ebene der in ihren Grundlagen näher erläuterten Funktionsweise von Computerartefakten, die jeweils trotz aller Fortschritte in der Softwaretechnik außerordentlich hohen intellektuellen Aufwand für Formalisierung und Modellierung der aufgabenspezifischen Prozesse kooperativer Wissensarbeit erfordern, in denen sie eingesetzt werden sollen (wobei Versuche, diese durch adaptive Systeme zu umgehen (s.o. (5)), noch größere Probleme aufwerfen können). Da zudem mit Computerartefakten als Medien der Kooperation tiefgreifend und selbstbezüglich in soziale Praktiken einer Organisation interveniert wird, sind diese Spezifikationsleistungen nur partizipativ in enger Kooperation zwischen Entwicklern und Nutzern zu erbringen. Letztlich laufen diese Anstrengungen auf eine integrierte Organisationsentwicklung mittels Computerartefakten hinaus. Darüber hinaus zeigt sich immer wieder, dass bei einzelnen eng begrenzten Aufgaben kognitiver Arbeit mögliche Einsparungen durch häufig anzutreffende Ausweitung der Funktionalität und Komplexität der Systeme und damit verbundene ›Rebound‹-Effekte wieder (über-)kompensiert werden.

(8) Die Gestaltung und Aneignung programmgesteuerter Verarbeitung bedeutungsloser Signale in Computern im Zusammenspiel mit sinngebenden sozialen Praktiken in Organisationen wird im Beitrag: *»›Informatik‹ – eine Wissenschaft auf Abwegen«* noch weiter vertieft analysiert. Er ist erschienen im 2022 von Gerhard Banse und Klaus Fuchs-Kittowski herausgegebenen Sitzungsbericht 150/151der Leibniz-Sozietät zur Tagung »*Cyberscience – Wissenschaftsforschung und Informatik. Digitale Medien und die Zukunft der Kultur wissenschaftlicher Tätigkeit«*.

In diesem 8. Beitrag werden die unzureichend erscheinende theoretische Reflexion des eigentlichen Gegenstands der ›Informatik‹ als Wissenschaftsdisziplin (treffender: Computing Science) und die irreführende Verwendung fragwürdiger Grundbegriffe wie ›Information‹ als mitverantwortlich für die Schwierigkeiten im praktischen Umgang mit Computertechnik kritisiert. Als begriffliche Alternative wird der von Peirce (1983) elaborierte triadische Zeichenbegriff als Grundlage theoretischer Reflexionen präsentiert sowie dessen Sinn und Nutzen begründet, um die angesprochenen Schwierigkeiten besser verstehen und bewältigen zu können.

(9) Gewissermaßen als bilanzierender Abschluss der mit diesen Beiträgen geleisteten kritischen Einschätzung der Potenziale und Herausforderungen konnektionistischer »KI«-Systeme wird unter der Überschrift »*Die Illusionsfabrik der*

›KI‹-Narrative« ein kurzer, eher populärwissenschaftlicher Beitrag präsentiert, der 2022 in einem »KI«-Themenheft der Zeitschrift FIfF Kommunikation publiziert wurde. Darin werden zur Einschätzung wesentliche Aspekte dieser Technik resümierend dargestellt.

5 Epilog: Der ruhmlose Sturz der »KI«-Denkmäler

Ein Großteil der medialen Propaganda um »KI« nährt sich von einer Reihe von einzelnen großartig als Durchbrüche gefeierten und auch als solchen inszenierten Erfolgserzählungen über »KI«-Leuchtturmprojekte. Sie erregen dann – ganz im Sinne der Hypothesen der »promising technology« – jeweils große öffentliche Aufmerksamkeit und werden entsprechend medial breit ausgewalzt, wie wir es gerade wieder mit ChatGPT erleben. In den letzten drei Dekaden seit Mitte der 1990er Jahre waren dies zuvor

- das Schach-Weltmeisterprogramm »Deep Blue« von IBM im Jahre 1996,
- die Sprach- und Wissensverarbeitungs-Software Watson von IBM, Sieger bei der US Quiz Show »Jeopardy!« 2011,
- das Programm AlphaGo Zero von Deep Mind, das 2016 den weltbesten Go-Spieler besiegte.

Auch die Wiederkehr solcher Leuchtturmprojekte ist ein durchgängiges Muster, das sich durch die ganze »KI«-Entwicklung der Computertechnik zieht. Daher ist es von einigem Erkenntniswert, sich den jeweils weiteren Verlauf der Entwicklung der Systeme genauer anzusehen.

Schachprogramme »Deep Blue« & Co.: Das von IBM entwickelte Schachprogramm »Deep Blue« wurde nach dem Wettbewerb nicht mehr weiterentwickelt. An seiner Stelle wurden aber mit anderen Programmen vergleichbarer Spielstärke (z.B. Hydra) weiterhin Experimente durchgeführt; insbesondere wurden neuartige Tourniere mit menschlichen Spielern aufgesetzt, bei denen es diesen erlaubt war, eigene Computer als Hilfsmittel zu nutzen. Der damals von Deep Blue besiegte Weltmeister Garri Kasparov sah sich auf dieser Basis zur Revanche herausgefordert. Mithilfe eines handelsüblichen PC, den er aufgrund seiner Erfahrung mit nützlichen Hilfsmitteln zur Analyse alternativer Spielzüge ausgerüstet hatte, gelang ihm ein nachhaltiges Comeback, das er in einem Artikel in der New York Book Review ausführlich reflektiert und kommentiert (Kasparov, G.: The Chess Master and the Computer, The New York Book Review, Feb. 11, 2010):

> »*Moravec's Paradox, in chess, as in so many things, what computers are good at is where humans are weak, and vice versa. This gave me an idea for an experiment. What if instead of human versus machine we played as partners? My brainchild saw the light of day in a match in 1998 in León, Spain, and we called it ›Advanced Chess‹. Each player had a PC at hand running the chess software of his choice during the game. The idea was to create the highest level of chess ever played, a synthesis of the best of man and machine. […]*

> *The computer could project the consequences of each move we considered, pointing out possible outcomes and countermoves we might otherwise have missed. With that taken care of for us, we could concentrate on strategic planning instead of spending so much time on calculations. Human creativity was even more paramount under these conditions. […]*
> *My advantage in calculating tactics had been nullified by the machine. […] The teams of human plus machine dominated even the strongest computers. The chess machine Hydra, which is a chess-specific supercomputer like Deep Blue, was no match for a strong human player using a relatively weak laptop. Human strategic guidance combined with the tactical acuity of a computer was overwhelming.«*

Es ist dies ein besonders eindrückliches Beispiel für die praxistheoretisch angeleitete IA-Strategie der Mensch-Computer-Interaktion, das Kasparov aufgrund seiner Analyse so bewertet. *»Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process. […] Where so many of these investigations fail … is by not recognizing the importance of the process of learning and playing chess.«*

Die schillernde Existenz von IBM Watson: Nach dem spektakulären Sieg in der Quiz Show »Jeopardy!« hat IBM für die wirtschaftliche Verwertung seines sehr aufwendig für Hochleistungs-Hardware entwickelten Programmsystems Watson zur Sprach- und Wissensverarbeitung 2015 einen eigenen Geschäftsbereich gegründet, der in verschiedenen Anwendungsfeldern (Krebsforschung und andere Gesundheitsdienste, (Rück-)Versicherungen u.a.) auf Basis von Watson hochgradig wissensbasierte Dienstleistungen anbieten sollte. Das hat aber nie wirklich und nachhaltig wirtschaftlich profitable Früchte getragen.

Nicht nur müssen für die verschiedenen Anwendungsfelder jeweils neue Wissensbasen und Verfahren der Hypothesenbildung aufgebaut werden, oftmals bleibt wie soft auch die Qualität der zum Aufbau der Wissensbasen verwendeten Daten fragwürdig oder schwer zu beurteilen. Dementsprechend lässt die Validität der durch das System ermittelten Ergebnisse häufig sehr zu wünschen übrig. Letztendes ist der Einsatz von Watson aufgrund dessen in verschiedenen Krebsforschungszentren (u.a. Houston, Heidelberg, Kopenhagen) gescheitert und die Sparte Watson Health wurde 2022 an die Private Equity Group Franzisco Partners verkauft, andere Reste in den expandierenden Bereich der IBM Cloud Services integriert. Offenbar ist menschliche Urteilskraft doch nicht so leicht zu ersetzen (Balzter, S.: Im Krankenhaus fällt die Wunderwaffe durch, FAZ 03.06.2018; IBM is selling off its Watson Health assets, NYT Jan. 21, 2022).

Der Thronsturz von AlphaGo: Im März 2016 gelang im Strategiespiel Go Google Alphabet mit seinem »KI«-Programm AlphaGo der publikumswirksame Sieg gegen den koreanischen Spitzenspieler Lee Sedol. Damit schien der *homo sapiens* einmal mehr durch eine aufwendig konstruierte algorithmische Maschine entthront. Fast genau sechs Jahre danach ist es nun dem US-Amerikaner Kellin Pelrine gelungen, zwei vergleichbar leistungsfähige »KI«-Systeme ohne direkte

Hilfe eines Computers zu besiegen, obwohl er selbst lediglich auf dem Niveau eines gehobenem Amateurs spielt (die Alphabet-Entwicklung wurde mit der Version AlphaGo Zero 2017 eingestellt). Dabei nutzte er eine Schwäche der »KI«-Programme mit der List, deren Spielweise auf einem Teil des Spielfelds quasi zu umzingeln, während sie mit Zügen in einem anderen Teil abgelenkt wird (Holland, M.: Sechs Jahre nach AlphaGo: Mensch besiegt erneut »zuverlässig« stärkste Go-KIs, heise online 20.02.2023).

So triumphiert am Ende, wie dieser doch eher ruhmlose Sturz der einst medial so hoch gepriesenen »KI«-Denkmäler von ihren hohen Sockeln bezeugt, vor allem eins: Die natürliche Intelligenz der aristotelischen praktischen Vernunft. Ihre ungebrochenen und uneinholbaren Potenziale tragen – wie der Igel im Wettlauf mit dem Hasen – schlussendlich den Sieg davon. Die ihr entspringende intuitive Einsichtsfähigkeit und analytische Kompetenz, ihre Fähigkeit zu Reflexion und kreativer Vorstellung, kurzum: ihre natürliche Intelligenz erweisen sich als nicht hintergehbar und nicht zu ersetzen. Warum muss eigentlich eingedenk dessen immer wieder versucht werden, mit Petaflops an Hardware-Leistung den aufwendigen und kostspieligen Nachweis zu führen, dass eine algorithmisch konstruierte Maschine umfangreicher und schneller rechnen kann als der *homo sapiens*, der sie doch aufgrund seiner praktischen Vernunft eigens zu diesem Zweck erfunden hat? Warum nur unternimmt er immer wieder den untauglichen Versuch, seine natürliche Intelligenz mittels Computer-Artefakten aufklären zu wollen, deren Konstruktion diese Intelligenz schon voraussetzt? Warum scheint es vielen Menschen partout nicht gelingen zu wollen, zu der grundlegenden Einsicht des berühmten Einsteinschen Chiasmus zu gelangen: *»Nicht alles, was zählt, kann gezählt werden, und nicht alles, was gezählt werden kann, zählt«*?

6 Literatur

Austin, J. L. (1962): How to Do Things With Words, Cambridge (MA): Harvard University Press

Bateson, G. (1985): Ökologie des Geistes. Anthropologische, psychologische, biologische und epistemologische Perspektiven, Frankfurt/M: Suhrkamp

Bender, E. M.; Gebru, T.; McMillan-Major, A. & Shmitchell, S. (2021): On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3-10, 2021, Virtual Event, Canada, New York (NY): ACM

Berkeley, E. C. (1949): Giant Brains or Machines That Think, New York: Science Editions Inc.

Bostrom, N. (2016): Superintelligenz. Szenarien einer kommenden Revolution. Berlin: Suhrkamp

Brödner, P. (1997): Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen, Berlin: edition sigma

Church, A. (1941): The Calculi of Lambda-Conversion, Princeton: Princeton University Press

Dreyfus, H. L. (1979): What Computers Can't Do. The Limits of Artificial Intelligence, New York: Harper & Row

Dreyfus, H. L.; Dreyfus, S. E. (1987): Künstliche Intelligenz. Von den Grenzen der Denkmaschine und dem Wert der Intuition, Reinbek: Rowohlt

↑Inhalt↑

Eco, U. (1994): Die Suche nach der vollkommenen Sprache, Frankfurt/M: Büchergilde Gutenberg

Eco, U. (1991): Semiotik. Entwurf einer Theorie der Zeichen, München: Wilhelm Fink Verlag

Frege, G. (1879): Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens, Halle: Verlag von Louis Nebert

Falkenberg, E. D. et al. (1998): A Framework of Information Systems Concepts. The FRISCO Report, IFIP,

Giddens, A. (1988): Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung, Frankfurt/M: Campus

Gödel, K. (1931): Über formal unentscheidbare Sätze der *principia mathematica* und verwandter Systeme I, Monatshefte für Mathematik und Physik 38, 173-198

Hirsch-Kreinsen, H. (2023): Artificial Intelligence: A »Promising Technology«, AI & Society, published online 18.01.2023; ausführlicher auch: Das Versprechen der Künstlichen Intelligenz. Gesellschaftliche Dynamik einer Schlüsseltechnologie, Frankfurt/M New York: Campus

Hoffmann, D. W. (2018): Grenzen der Mathematik, 3. Aufl., Berlin Heidelberg: Springer Spektrum

Hofstadter, D. R. (1979): Gödel, Escher, Bach. An Eternel Golden Braid, New York: Vintage Books

Kleene, S. C. (1952): Introduction to Metamathematics, Amsterdam: North Holland

Kurzweil, R. (2006): The Singularity Is Near: When Humans Transcend Biology, New York: Penguin Group

McCarthy, J. et al. (1955): A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, http://jmc.stanford.edu/articles/dartmouth.html

McCarthy, J. (1960): Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I, CACM 3 (4), 184-195

McCulloch, W. S. & Pitts, W. A. (1943): A Logical Calculus of the Ideas Imminent in Nervous Activity, Bull. Math. Biophys. 5, 115-133

Nake, F. (2001): Das algorithmische Zeichen, in: Bauknecht, W.; Brauer, W.; Mück, T. (Hg.): Informatik 2001. Tagungsband der GI/OCG Jahrestagung, 736-742

Nake, F. (1992): Informatik und die Maschinisierung von Kopfarbeit, in: Wolfgang Coy et al. (Hg.): Sichtweisen der Informatik, Braunschweig Wiesbaden: Vieweg, 181-201

Newell, A. (1980): Physical Symbol Systems, Cognitive Science 4, 135-183

Nonaka, I. (1994): A Dynamic Theory of Organizational Knowledge Creation, Organization Science 5 (1), 14-37

Peirce, C. S. (1983): Phänomen und Logik der Zeichen, Frankfurt/M: Suhrkamp

Peirce, C. S. (1934): Collected Papers Vol. 5, Pragmatism and Pragmaticism, Cambridge (MA): Harvard University Press

Polanyi, M., 1985: Implizites Wissen, Frankfurt/M: Suhrkamp (engl. Original: The Tacit Dimension, Garden City (NY): Doubleday 1966)

Ropohl, G. (1991): Technologische Aufklärung. Beiträge zur Technikphilosophie, Frankfurt/ M: Suhrkamp

Ryle, G. (1987): Der Begriff des Geistes, Stuttgart: Reclam (engl. Original: The Concept of Mind, London: Hutchinson 1949)

Schon, D. A. (1983): The Reflective Practitioner: How Professionals Think in Action, New York: Basic Books

Searle, J. R. (2012): Wie wir die soziale Welt machen. Die Struktur der menschlichen Zivilisation, Berlin: Suhrkamp

Shannon, C. E. (1948): A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379–423 & 623–656

Taylor, C. (2017): Das sprachbegabte Tier. Grundzüge des menschlichen Sprachvermögens, Berlin: Suhrkamp

Teich, I. (2020): Meilensteine der Entwicklung Künstlicher Intelligenz, Informatik Spektrum 43: 276–284

Turing, A. M. (1936): On Computable Numbers. With an Application to the Entscheidungsproblem, in: Davis, M.(Ed.): The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems, and Computable Functions, New York 1965, 116-151

VDI (Hg.) (1991): Technikbewertung – Begriffe und Grundlagen. Erläuterungen und Hinweise zur VDI-Richtlinie 3780, VDI Report 15, Düsseldorf: VDI

Vaswani, A. et al. (2017): Attention Is All You Need, arXiv:1706.03762v5 [cs.CL] 6 Dec 2017

Volpert, W. (1999): Wie wir handeln – was wir können. Ein Disput als Einführung in die Handlungspsychologie, 2. überarbeitete und aktualisierte Auflage, Sottrum: artefact Verlag

Weizenbaum, J. (1977): Die Macht der Computer und die Ohnmacht der Vernunft, Frankfurt/M: Suhrkamp

Wing, J. M. (2006): Computational Thinking, CACM 49 (3), 33-35)

Winograd, T. & Flores, F. (1986): Understanding Computers and Cognition. A New Foundation for Design, Norwood: Ablex Publ.

Xin, R. (2016): word2vec. Parameter Learning Explained, arXiv: 1411.2738v4 [cs.CL] 5 Jun 2016

↑Inhalt↑

Coping with Descartes' Error in Information Systems

Abstract: Coming from Hubert Dreyfus' recent book »Retrieving Realism« (together with Charles Taylor), the paper presents embodied pre-conceptual perception and representational cognition as two contrasting perspectives on accessing the world. It further characterizes the different forms of knowledge emerging from these perspectives and how they dynamically relate to each other. Taking up the Peircean theory of signs and abductive reasoning as methods of discovery, computers are analysed as semiotic machines that formally model and objectify explicit knowledge about social practices and that can be embedded in the sign processes of thereby restructured practices. This practice theoretical perspective allows for both, understanding the limits of AI and pointing to options for productively combining the performance of »cognitive artifacts« with the tacit skills of knowledge workers.

Key words: Abductive reasoning, artificial intelligence, implicit experiential knowledge, explicit propositional knowledge, semiosis, semiotic machine.

1 Introduction: Taking up Dreyfus' legacy

In his most recent book »Retrieving Realism« (2015; together with Charles Taylor), Hubert Dreyfus reminded us again of two fundamentally different thought traditions on how we get access to the world around us: the »contact« theoretical perspective of »being-in-the-world« which we make sense of through skillfully interacting and coping with it, on the one hand, and the representational perspective of conceptually mediated world recognition, on the other. While Dreyfus argues for the relevance of the first view particularly referring to Wittgenstein's (2009) »language-games«, Heidegger's (1962) ontology in »Being and Time«, and Merleau-Ponty's (1962) »phenomenology of perception«, the latter has dominantly influenced Western thinking since the days of Descartes' dualism separating mind from matter.

Both, the contact theoretical and the representational world view have been propagated for a long time in isolation and opposition with each other. Moreover, both correlate with two fundamentally different types of knowledge being characterized as implicit and embodied or experiential knowledge, on the one hand, as opposed to explicit and conceptual or propositional knowledge, on the other (cf. e.g. »knowing how« versus »knowing that« (Ryle 1949) or the »tacit dimensi-

on« (Polanyi 1966)). Both views can claim to be based on sound evidence, and it seems impossible to refute one or the other.

The long lasting controversy between both perspectives is, however, by far not an idle dispute among academics but rather of particular relevance for social practices dealing with complex computing machinery. It forms the epistemological background for designing, evaluating, and appropriating this specific semiotic type of machinery operating signs as distinct from energy or material transforming machinery operating with forces. In particular, the representational view seduces the creators of so called artificial »intelligence« (AI) or »smart« machines to claim to either mimic (»weak AI«, as assessed by the »Turing test« (Turing 1950)) or even represent human intelligence (»cognition is computation« (Pylyshyn 1984), »strong AI«). If all our knowledge would lastly be explicit and propositional, it in fact might also be represented by algorithms and implemented as a computing machine, either by logical or functional programming or by machine »learning«.

Against this background, the paper will shortly present both epistemological perspectives in order to investigate their actual dynamic interrelationship: Regarding humans primarily as bodily and socially embedded actors dealing with the world around them, they are seen as being able to primarily gain an immediate, pre-reflexive perception of the world through their interaction driven by needs. It is implicit, embodied knowledge, a tacit skill private to the actors, and it normally is sufficient to cope with the world and the things »present-at-hand«. Driven by curiosity to explore the world or when actions lead to irritation, surprise or failure it might become necessary to reflect on the specific situation by conceptually reconstructing the underlying actions and processes. In this reflective stance, by conceptualizing, both by self-observation or observation by others – i.e. by distinction and designation –, explicit or propositional knowledge about practices is being created.

This conceptual knowledge about practices is secondary and representational, hence, it is necessarily incomplete and partial, and it is error-prone (rightly evoking Cartesian doubts). Moreover, this theoretical knowledge needs to be appropriated again for situated practical use to become actually effective. Appropriation is work to make explicit knowledge (or knowledge-based artifacts) work, an effort which, on his part, enriches the implicit or tacit practical knowledge. With this dynamic interrelationship in mind, Descartes' error does not consist in expressing his doubts about the certainty of propositional knowledge, but rather in regarding this representational cognition as the primary and only way of perceiving the world.

The practice theoretical perspective with its focus on the dynamics of explicating successful social practices as conceptual knowledge and of appropriating such knowledge as enhanced tacit knowledge and practical skill is of particular relevance to understand the persisting problems with conventional approaches to

designing and appropriating complex computing machinery for use in organizations and the productivity paradox associated with them. It points to the need for understanding the semiotic nature of computing systems and the limits of their performance.

To this end, the paper starts with shortly characterising what it means that humans are living organisms being-in-the-world and what the pre-representational perception of the world is about. This lays the ground to analyse the dynamic relationship of tacit and codified knowledge. As a consequence of this distinction and relationship, the question arises how new ideas can be generated and introduced in codified knowledge leading to the logical form of abductive reasoning. With these basic cognitions in mind, computers are characterised as semiotic machines and the problematic relationship between humans and allegedly »smart« machines is analysed and assessed. Finally, conclusions about the potentials of human action competence and the limits of computer performance are drawn.

2 Being-in-the-world

In the newly flaming up AI discourse, spectacular technical demonstrations cause a big stir and confusion again, although they are build on well-known old conceptions of cognition. There is a continued prevailing tendency to compare or even equalize autonomous intentional acting of humans with the other-directed functional behaviour of machines. It particularly tends to confuse functional isomorphism with identity, equality of form with equality of substance. With respect to these tendencies, it appears necessary to recall some very basic facts about what it means to be a living human interacting with the world.

To begin with, humans are, like animals, first of all, natural beings, existing in a natural world around them which keeps them alive through metabolism. As living organisms, coming into being through the self-organising dynamics of »autopoiesis« forming their ›inner‹ nature, they interact with the ›outer‹ nature surrounding them in need of being nurtured by digestible natural products at reach. While the sensory-motor capacities of their organism form the interface between inner and outer nature, the quality of this dependence can be sensed and felt as well-being or illness (Maturana & Varela 1992). Moreover, humans are, as ›social animals‹, born into a community of others and need to be looked after during a long phase of adolescence and socialisation. As individual members of the community, they collectively make provisions for their living.

The human organism's inner nature is, however, not only passively dependent on a favourable outer nature, but has, by evolution, also some reflective and productive powers at its disposal to actively get what it needs. Humans are, other than animals, equipped with the capacity of knowing about and of reflecting upon their being-in-the-world and to develop consciousness (an irrefutable fact the emergence of which we cannot fully explain scientifically so far, though). In con-

trast to the Cartesean conception of a rational mind being independent of bodily experience and feeling, human perception and cognition rely on manifold interactions between the brain and the body of which it is an organic part (Damasio 1994). As a natural evolutionary heritage, this active and deliberate »being-in-the-world«, together with the perceptual and motor capabilities it produces and maintains, provides humans with the necessary skills to survive (originally living as collectively acting »hunters and gatherers«).

Against this background, Hubert Dreyfus and Charles Taylor (2015) remind us again of the primacy of embodiment emphasizing the living body as the primary site of experiencing the world and of the fundamental role immediate perception plays in engaging with it. Referring to Heidegger's ontology in »Being and Time« (1962), and Merleau-Ponty's »Phenomenology of Perception« (1962), he insists that the world and the human body with its rich sensory-motor functions of perceiving the world are intricately intertwined. This interaction between inner and outer nature triggering their mutual momentum (also known as »structural coupling«; Maturana & Varela 1992), enables the body to feel what it needs and to find what the world offers to still the needs; it amounts to continuous skilful coping with the emerging situations of actively being-in-the-world. The active body and that which it perceives cannot be disentangled, while its saturation and well-being indicate success in this relationship.

Reflecting on indications of crisis in natural sciences, the late Edmund Husserl had already emphasised the significance of what he called the primarily experienced »life-world as the forgotten meaning-fundament of natural science: […] we must note something of the highest importance that occurred even as early as Galileo: the surreptitious substitution of the mathematically substructed world of identities for the only real world, the one that is actually given through perception, that is ever experienced and experienceable – our every day life-world« (Hussserl 1970: 48f). He thus laid open the root of taking a symbolic »garb of ideas« for objective reality confusing the method or model with actuality.

The conception of a non-mediated, pre-representational, and pre-reflective activity driven by »motor intentionality« (Dreyfus & Taylor 2015) leads to the view of an embodied and socially embedded actor grappling with the physical and socio-cultural world around him. In this perspective, the pre-representational coping with the world is not being produced by the actor alone, but in co-construction with the world by interacting with what it offers, with the »things-present-at-hand« (Heidegger 1962). The phenomenal »thing« is, in this view, not the unchanging object of the natural sciences, but a correlate of our body and its sensory-motor functions. According to this »contact theory« of human existence, the thing's experience and meaning result from the interaction, they are not internal to the actor, but lie in the interspace of dealing with it. Driven by curiosity and intentionality, the actor thus attains an intuitive percipience of the whole situation in

which he is acting and, in particular, of the meaning things have in situated use without being conspicuous.

The phenomenological »contact« theoretical world view emphasizing these facts has been brought forward against the representational view of conceptual cognition, against Descartes' cogito prevailing in the Western world to which it indeed stands in stark contrast (Damasio 1994, Dreyfus 2002, Ryle 1949). Both world views have long been regarded as opposing or even excluding each other. They may, however, also be looked at as mutually completive views taking into account the additional fact that humans are able to awake to their being-in-the-world. Consciousness is a higher level of percipience transcending feeling, experience, will, and skills; it comprises, beyond immediate perception (as outlined), at least two more basic capacities: first, to perceive something and to be aware of perceiving it at the same time, and second, to perceive it as something, i.e. to subsume it under a concept. Consciousness thus enables humans to perceive and act (intuitively and immediately) on the one hand and to reflect on what they perceive and do (conceptually and analytically) on the other.

Individuals become intentional actors by virtue of observing how others are mirroring their acting. Based on this basic reflective capacity, humans are able to experience as well as to observe themselves as acting intentionally and, in particular, cooperatively, i.e. to conceive themselves as social actors, in other words: to say »I«. By becoming conscious of their position in these natural and social relationships, humans are enabled – and forced at the same time – to jointly take care of their lives, to take their specifically human needs as guideline for their deliberate and active intervention in or rearrangement of their living conditions, in other words: they become able to work (this capacity being fully developed during the neolithic revolution). Their bodily existence within these natural relationships forming an objective material precondition for human living enables humans to discover the creative or productive powers slumbering within them in order to take possession of the outer nature, to extend its potential for improving their living conditions with care. This inescapable fact also confronts humans with the necessity, however, to reflect on the risks such interventions may generate for maintaining the living conditions.

In the course of their conscious engagement with the physical and socio-cultural world, humans have developed two basic productive capacities: dealing with tools (as technical acting) and dealing with signs (as signifying acting). Both capacities are based on the formation of concepts (representations) as abstract generalised experiences and their objectification that can be shared; they thus initiate a cultural evolution based on and superimposing the original natural relationships. Tools are purposefully shaped artifacts complementing and augmenting either sensory or motor organs of the body for more effective observation of or intervention in the outer nature, while signs and significations are socially generated enti-

ties enabling humans to reflect on what they are doing, to mentally perform trials, to communicate with each other, and to coordinate joint actions.

Tools such as hand-axes, ploughs, or knifes are positioned as intermediaries between the inner and outer nature tying the needs and capacities on the part of the acting subject with adequately constructed properties on the part of the tool. For appropriately mediating between the two, tools need to be designed and mastered according to the purpose pursued within the constraints set by formability of nature and usability of things. For effective practical use, tools need to be sufficiently appropriated, internalised and wielded; only then the acting subject can merge with the intended object in specifically skilful actions.

Physical signs such as gestures, speech sounds, or written characters, on the other hand, serve as representations referring for somebody in a certain respect to something else as denoted object, i.e. having a meaning for somebody in a specific action context. Besides this representative function, signs further can be used as cognitive means in the sense that all cognition is mediated by signs (in contrast to immediate percipience as outlined above). In particular, they can be used to represent absent or even purely thought-of objects, to describe plans for instance. Signs do not exist per se, but are always part of a process of signification, or »sign process« (Peirce 1903) as a social fact. Signification or sign processes (»semiosis«) depend on collective intentionality of the actors enabling communication and cooperation in a social collective (as has been approved only recently; cf. Searle 2010, Tomasello 2008; 2014). Signs are our »windows« to the world through which we grasp or conceive the signification of certain aspects of world; using signs implies that there is no world without the meaning of »world«.

3 The dynamics of »knowing how« and »knowing that«

The contact theoretical, pre-representational as well as the representational views on perception (or the two levels they refer to, respectively) also correlate with two corresponding, fundamentally different types of knowledge: implicit, practical or tacit knowledge (Polanyi 1966) on the one hand, and explicit, theoretical or propositional knowledge on the other. This appears to be a very basic distinction which is referred to in a whole bundle of literature in a similar way (Dreyfus & Dreyfus 1986, Giddens 1984, Göranzon & Josefson 1988, Nonaka 1994, Polanyi 1966, Ryle 1949, Varela et al. 1991). While human activities and related perceptions produce, according to the pre-representational world view, a wealth of embodied experiences, skills, and tacit knowledge, observing and conceptually reflecting on social practices of acting collectively, associated with the representational world view, lead to explicit or propositional knowledge as objectifications.

Tacit knowledge (implicit »knowing how«; Polanyi 1966, Ryle 1949) is the practical human action competence emerging from interacting with the world and comprising reflective, operational, and co-operational capabilities, skills, and ex-

periences the human living body develops during lifetime. They enable effective acting in specific situations, particularly judging complex relationships and coping with uncertainty, in order to achieve needs and interests. They amount to successful »situated action« (Suchman 1987). Knowing how is the result of our intentional relationship to the world. It embraces the capacity to perceive and capture a situation as a whole, not as a constellation of separated parts and properties, as well as procedural routines for appropriately continued action. This action competence is mostly unconscious (»tacit«); it grows with experience and by appropriating and internalising explicit knowledge as well as technical artifacts for effective practical use.

Propositional knowledge (explicit »knowing that«; Ryle 1949), in contrast, consists in making experiences conscious through reflection and conceptualisation. By means of appropriate concepts, certain aspects of social practices in the world are being ordered and partially explicated in propositions *about* the practices. Hence, explicit knowledge is always partial, perspective, and limited: It is limited with respect to capabilities as experts know more than they can tell (Dreyfus & Dreyfus 1986, Polanyi 1966). It also is limited logically as concepts and propositions comprise only what they stand for (excluding everything else). Explicit knowledge takes the form of theories, i.e. systems of consistently related concepts and propositions. Theories explain how something works (comprehending); they are self-referentially closed and need to be appropriated for effective practical use (real problem solving). The limitations are the price to be paid for conceiving aspects of our being-in-the-world, for obtaining abstract and generalised knowledge about it.

Both types of knowledge do not exist independently in isolation, but rather are intertwined in such a way that they mutually produce each other under certain conditions. Experience and tacit »knowing how« (Ryle 1949) as a result of our bodily existence is the primary basis of all acting and perceiving; as such it establishes a social practice and it is disposable at any time (although it may eventually turn out not to be sufficient). The intuitive and antecedent actions and perceptions generating experience normally express themselves as a continuing successful practice: »A successful practice precedes its own theory« (Ryle 1949: 30). Explicit »knowing that«, in contrast, is secondary and limited; it is, however, objectified (externalised) and represented by signs and can, therefore, be cumulated and communicated to others. It emerges by reflecting on and explicating an existing social practice, and it needs to be appropriated and internalised on its part in order to meaningfully affect a social practice, though. As it roots in abductive reasoning, it is prone to fail, and Descartes was right to maintain his fundamental doubts about the certainty of cognition in the form of propositional knowledge. His deep error and that of his followers up to the AI propagandists is, however, to regard this type of knowledge as the only one, thus overemphasizing its relevance for

effective acting in the world, while ignoring the tacit dimension of knowing and the phenomenology of pre-representational perception (Dreyfus 2002).

Taking a practice theoretical perspective (Reckwitz 2002) on this dynamic relationship by which new knowledge is created (Nonaka 1994), the actors of a social practice are, in the flow of continued collective activity, primarily performing internalized, embodied routines as a product of habituation. Performing a social practice may also comprise dealing with things such as technical artifacts through which the effects of acting are augmented, while new action routines are formed and embodied in use. These internalized routines can analytically be differentiated into acts of signification, domination, and legitimation; altogether they enable and constrain further acting as taken for granted according to the practice theoretical view of e.g. structuration theory (Giddens 1984; with particular regard to computer artifacts: Rohde et al. 2016; cf. the inner loop in figure 1).

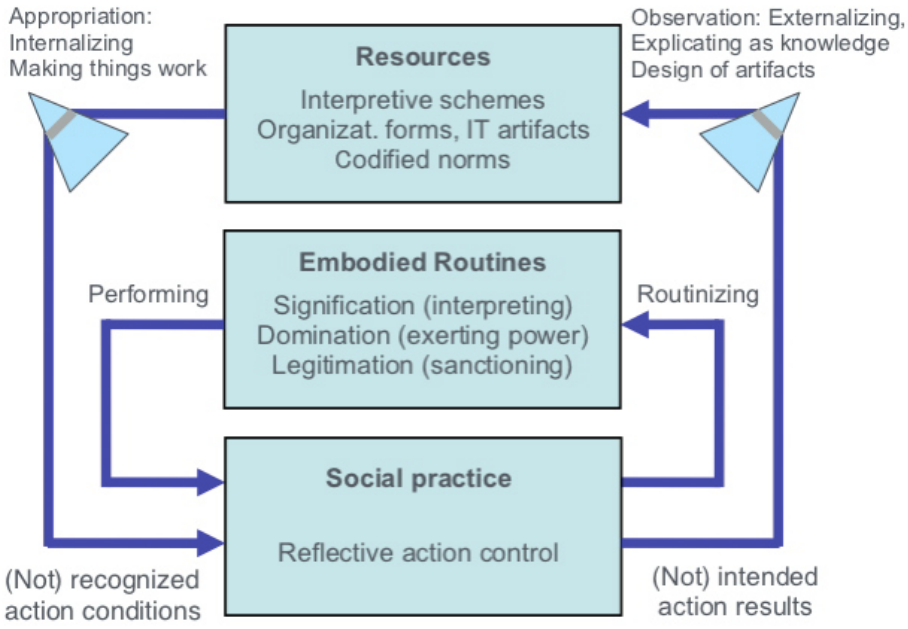


Fig. 1: Structuring a social practice (own representation)

Both, performing routines or appropriating resources constitute a set of action conditions (recognized or not) for further acting in ongoing social practices, the effects of which deliver results (intended or not) that (re-)structure routines. In the course of this continuous collective action flow, moments of irritation or surprise may occur, where things become conspicuous for whatever reason (possibly because established routines fail) and attract specific attention. Such problems lead to a situation in which things normally taken for granted lose their »objectivity«, since objectivity is not naturally given, but ascribed through shared signification. The experienced disorientation in such action crises not only relates to the object, but also concerns the social actors themselves. It initiates reflection and search

processes in order to regain the capacity to act appropriately (as e.g. treated with notions of »break-down« and »reflection-in-action« by Schon 1983).

Remedy would normally be achieved by reflecting on and conceptualizing routinized action patterns in explicit terms according to the logic of abduction (Peirce 1935), i.e. by forming appropriate conceptual hypotheses which provide a ›best fit‹ with previous experience and knowledge, seeking to explain and to transcend the problematic situation. Actors are able, in this way, to reframe their knowledge and to test their new understanding, to internalize it if proven to be effective, and thus to regain the capacity for effective routine action. This capacity then also includes the ability to anticipate the possible functions and properties of artifacts, learned from previous actions.

Material resources, together with the appropriated routines to handle them, like other internalized action routines (signification, domination, and legitimation), constitute social structures that enable and, at the same time, constrain collective acting (»duality of social structure«; Giddens 1984). By making sense of the internalised resources »present-at-hand« through interpretation (signification), by sanctioning actions according to codified norms (legitimation), by influencing other actors through administrative resources or by shaping activities through the use of technical artifacts (like e.g. software functions; domination), they both continuously (re-)create routines, and eventually develop further resources, that constrain the scope for future action, interaction and negotiation. The more material resources are adjusted to the action context and the more appropriately they are interpreted and appropriated (or »encarnated«) for practical use, the more effective and efficient the social practices will be (Rohde et al. 2016).

In sum, humans act with the artifacts at hand by virtue of the meaning they attribute to the artifact's functions and the results they produce. By making sense of and effectively making the artifacts' functions work in use, specific regularities and use patterns emerge, which become internalised as new routines. Through recurrent interaction with the artifacts at hand, certain of the artifact's functions or properties thus become implicated in an ongoing process of structuration in which rules and routines of use emerge. The resulting recurrent social practice produces and reproduces a particular social structure of artifact use.

With respect to the use of internalised artifacts as analysed by Merleau-Ponty (1962) – or taking up similar results from the alternative perspective of activity theory (Engeström et al. 1999, Leontiev 1978) –, two different classes of artifacts need to be distinguished regarding capacities and skills involved in using appropriated artifacts. Internalised artifacts that mediate motor skills makes them a part of the body schema such that they become a medium through which motor skills are expressed. This mediation of motor skills may either be expressed through artifacts serving as tools for physically interacting with the environment requiring wielding skills in handling the tools. Or artifacts serve as appendages to the body by which it moves through the environment requiring navigational skills. For

most artifacts used in this way, the perceptual functions are subordinate to their motor functions. By being appropriated and incorporated into the body schema, the artifacts become part of the bodily space (»space of situation«; Merleau-Ponty 1962), thus becoming an integral part of the motor or perceptual skill repertoire. In any case, embodied artifacts serve as media through which motor or perceptual functions are expressed, and they typically enhance or extend the performance or potential reach of perceptual and motor skills.

Besides those embodied artifacts mediating and extending motor and perceptual skills, there is another totally different class called »cognitive artifacts« (Norman 1993: 47ff) which are designed to manipulate, store, or retrieve physical signs representing socially relevant information to be generated by their users (e.g. computers, calculators, forms, or books). They support and eventually extend cognitive abilities, such as thought or reflection, memory, problem solving, or language use. Cognitive abilities or skills are grounded in, but not directly reducible to, sensory-motor skills, as the embodied use of such media conveying representations in the form of signs require additional capabilities to appropriately interpret and make sense of the signs in a specific situation that go beyond perception and manipulation.

4 Abductive reasoning

According to the outer loop in figure 1, explicating experience as propositional knowledge or designing a technical artifact's functions based on such knowledge as well as, conversely, appropriating such knowledge or those functions for effective practical use are both creative actions requiring tacit skills for their part. These creative skills are logically based on *abductive* reasoning, a a third type of inference (besides induction and deduction) launched by Charles S. Peirce addressing logical problems of discovery. Since the genesis of a hypothesis was an open question in scientific inquiry so far, he introduced abduction as the only logical operation dealing with a new idea transcending existing cognition. One basic way of formulating abduction is the following (Peirce, CP 5.189):

> »The surprising fact, C, is observed.
> But if H [an explanatory hypothesis] were true, C would be a matter of course. Hence, there is reason to suspect that H is true.«

Abduction can thus be understood as a mode of inference in search of explanations for puzzling or anomalous phenomena or events. It allows to more comprehensively understand the process of scientific inquiry as a three step procedure of abduction (in search of an appropriate hypothesis), deduction (for deriving facts to test), and induction (to interpret the test results).

The abductive mode of inference can, however, be criticised with at least two respects: First, because it might be too permissive to be of much use as it seems to permit inferences to all sorts of wild hypotheses. Second, because it still not really

addresses the genesis of the hypothesis, a creative act that clearly transcends logical reasoning. Both critical aspects point to the need for the researchers' implicit »knowledge of familiarity« (Göranzon & Josefson 1988: 17), their tacit skills in »seeing« similarities or analogies when looking at the new or puzzling situation. The skill of creating a suitable hypothesis based on analogy or ›likeness‹ without following explicit rules is an essential part of human intelligence. In fact, many historical cases of scientific discovery illustrate this creative moment of »eureka« experience. It often relies on a form of sign-based reasoning called »diagrammatic reasoning« by Peirce.

There are a multitude of cases to demonstrate diagrammatic reasoning. The method by which Thales of Milet has determined the height of a giant pyramid provides an illustrating example. Taking the experiential cognition that objects of different height throw shadows whose lengths stand in equal relationship as their heights (cf. figure 2) as abductive reason, he could measure the pyramid's shadow p and compare it to the more easily measurable shadow s of an upright stick of length S. The unknown height P of the pyramid can thus be calculated from easily measurable values.

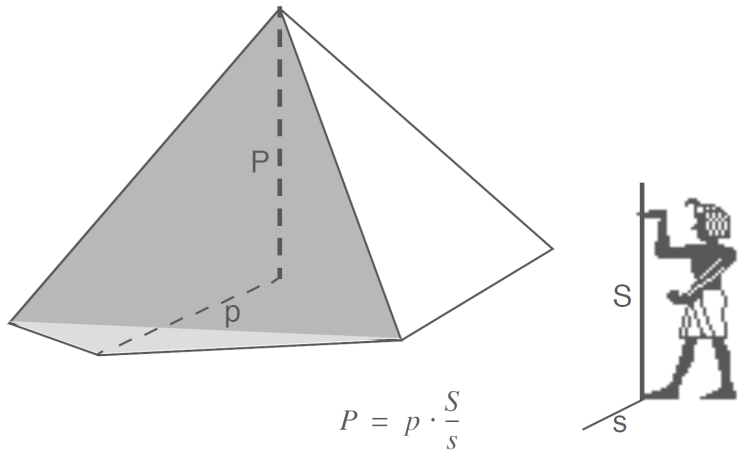


Fig. 2: *Determination of a pyramid's height (own representation)*

Another specifically illustrative example for diagrammatic thinking early in history is the proof of the famous »Pythagoras theorem« which actually goes back in time as far as at least four thousand years (long before Pythagoras lived): In the old fluvial cultures of Mesopotamia and Egypt with its high agricultural surplus product, there were a strong need as well as elaborated practices for disposing and surveying field areas. The basic tool for this was a closed loop rope with 12 (= 3 + 4 + 5) equidistant marks on it to create a right angle at any place according to the Pythagoras theorem (e.g. 9 + 16 = 25):

A triangle is right-angled <=> $a^2 + b^2 = c^2$.

A theorem proof (one among many possible others) builds on the widespread social practices and experiences in dealing with geometric areas at the time. Dia-

grammatic thinking playing with geometric areas immediately leads, starting from the lower left rectangular triangle, to the geometric configuration shown in figure 3 where the big square with the area $(a + b)^2$ is composed of the square c^2 plus four right-angled triangles with the area $(a\ b)/2$ each. This results in the equations shown in figure 3.

$$(a + b)^2 = c^2 + 4\ (a\ b)/2$$
$$a^2 + 2\ a\ b + b^2 = c^2 + 2\ a\ b$$
$$a^2 + b^2 = c^2 \qquad \text{q e d.}$$

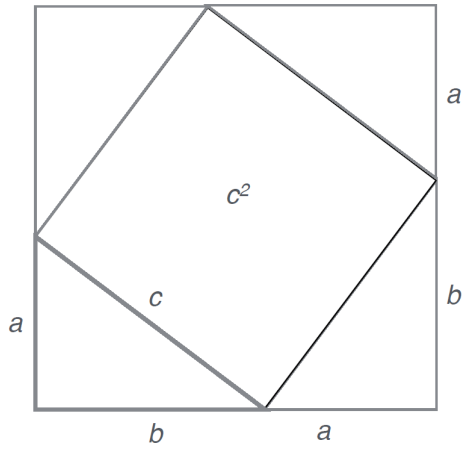


Fig. 3: *Proof of the Pythagoras theorem (own representation)*

5 Computers as semiotic machines

As indicated by the the wide-spread denomination »information technology«, computers have for a long time been regarded as »information processing« machines – cf. e.g. the Académie Française definition for »Informatique« (computer science) as »rational, in particular automatic, processing of information«. This view is, however, extremely misleading, as information is a concept belonging to the social world of signification, of assigning meaning to processes or events: Information, i.e. »any difference that makes a difference« (Bateson 1980: 250), solely originates from the activity of social actors interpreting such processes or events in the context of a social practice. As such information simply does not exist inside a computer system clearly operating in the physical world of causes and effects.

That which is actually operated in computers are physical signs or signals being processed simply following the functional instructions of an algorithm. Physical signals are thus being processed by computable functions (as specified by the conception of the Turing machine) – and nothing else. Hence, there is no mystery at all, the computing machine explicitly does not process information, and it does not, in any comprehendible sense, »know« what is the meaning of the signals processed, nor does it »know« what it is doing while it performs those functions.

Equally, Winograd & Flores (1986: 86 f), referring to signal processing, also emphasize: »One could describe the operations of a digital computer merely as a sequence of electrical impulses traveling through a complex net of electronic elements, without considering these impulses as symbols for anything.«

Signification is not a property of symbols, but rather is assigned through interpretation in the context of a social practice, construed in use. There seemingly is an unsurmountable gap between the social world of significations, of assigning meaning in sign processes, and the physical world of signal processing in computer systems. Both worlds can be brought together, however, by merging the technical computer functions with the social practices using them according to the practice theoretical perspective outlined. Such a *sociotechnical* system may be analysed in detail by taking up the triadic sign concept as elaborated by Peirce (1903). He looks at signs and processes of ascribing signification as a triadic relationship connecting three entities: a physical sign (signal or »representamen« *R*), the »object« *O* it denotes or refers to, and the »interpretant« *I* or meaning it assigns to this relationship (figure 4).

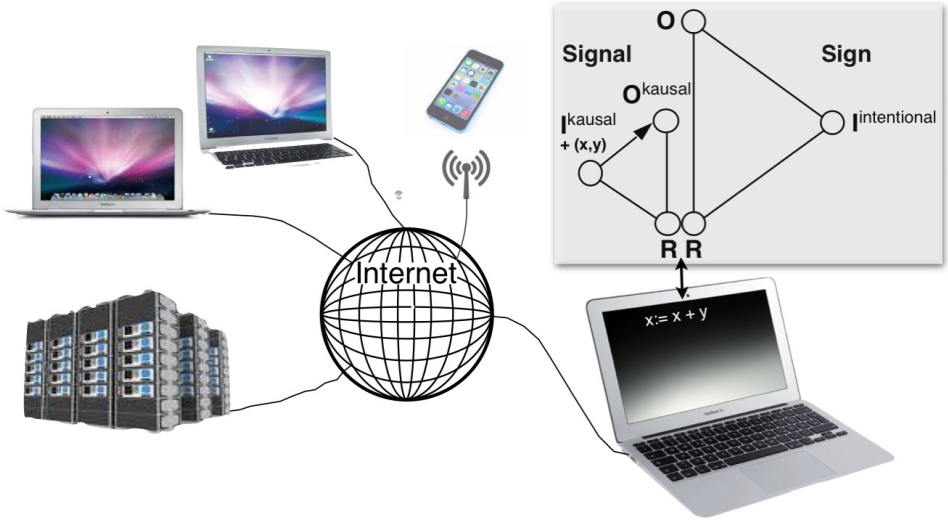


Fig. 4: *Algorithmic sign mediating signal and signification (own representation)*

In particular, this triadic sign concept allows to comprehend the algorithmically determined signal processing in computer systems as »degenerated« sign processes reduced to a dyadic relation without a »window to the world«, i.e. lacking the reference to an object of experience (its denotation). It only is a »*quasi*-semiosis« operating solely with physical signals as »*quasi*-signs« (Nöth 2002). This reduced *quasi*-semiosis can, however, be embedded in or merged with a complete semiosis or sign process of a social practice by means of a common »representamen«, i.e. a definitely coded physical signal made accessible to human senses (e.g. on a screen) which as such can be subject to interpretation in the context of a

social practice outside the computer (provided that the algorithmic functions in use are known).

This kind of semiotic analysis equally applies to the use of computers as »embedded systems«, as controllers of physical (or biological) processes in so called »cyber-physical systems«. In these cases, the systems or processes to be controlled need first to be sufficiently described and modelled in the form of sign-based heuristic or mathematical models that allow for designing appropriate control algorithms. By means of these control algorithms, the computer then directly operates relevant sensory signals for generating actor signals, solely in the form of a »*quasi*-semiosis«, in order to grant an intended automatic system or process behaviour.

The semiotic nature of computers may be illustrated by analysing sign processes in the social practice of computer-supported knowledge work referring to what has been specified as »algorithmic signs« (Nake & Grabowski 2001): The use of computers in this setting is based on two coupled sign processes interlinked by the same representamen R (irrespective of reversibly definite codification). While interacting with the computer, humans use triadic signs as input being meaningful to them in their social practice. Inside the computer system, however, these signs, being readable and meaningfully interpretable in the outside action context, are reduced to pure electronic signals as »quasi-signs«. The signals don't »know« any more for what they stand and what they mean. Rather, they are being processed through a program according to the completely determined instructions of the underlying algorithm. In Peircean notation, the algorithmic instructions in this sign process reduced to syntactical operations on signals take the role of an interpretant, however a »causal interpretant« Icausal that formally falls in one with the designated object Ocausal. As soon as the processed signals – in the example shown the add-operation +(x,y) – are accessible to the senses again e.g. on a screen, they can, provided that the underlying algorithmic functions are known, be interpreted in the context of the social practice using the computer system's functions (Brödner 2009; cf. figure 4). Inside the computer system, the signal processing is fully determined by semiconductor physics and formal logic, while outside the corresponding signals are taken as full signs and subject to intentional interpretation.

6 Human intelligence and »smart« machines

Taking up the view of humans as living organisms being-in-the-world again, a number of basic facts about the ontological status of humans, (semiotic) machines and their relationship can now be stated:

- While humans grow and literally »make« themselves through the self-organising dynamics of »autopoiesis« (Maturana & Varela 1992), by metabolism and conscious interaction with the world, computers as semiotic machines

are, like other artifacts, made for human purposes using explicit theoretical knowledge about the world.

- While humans act autonomously and intentionally in a self-determined (contingent) way, semiotic machines operate automatically with a causally determined behaviour.
- While humans are able, thanks to their skill of dealing with signs, to learn through reflection and insight, semiotic machines can at best adapt to environmental conditions controlled by algorithms.
- While humans possess implicit tacit knowledge, growing experience, situated judgment, and action competence which are expressed and, at the same time, variably reproduced through action (forming the core of working capacity), the behaviour of semiotic machines is controlled by algorithms based on a practice's formalised sign processes; the algorithmic functions need to be appropriated for becoming effectively used in a thereby changed practice, in particular for managing prevailing processes not modelled and formalised so far.
- While the physical or algorithmic functions of machines completely underly the purpose they are designed for, the purpose itself is set by collective intentionality of social actors and, hence, subject to their interests and world views: Technical artifacts are socially embedded.

From the very beginning, computers have been described by extremely misleading metaphors like »electronic brain«, »autonomous« or »self-organising behaviour«, »intelligent«, »smart«, or even »self-healing machines«, »machine learning«, or »neural nets« (now being omnipresent). In particular, »artificial intelligence« is a strongly mistaken attribution: The word »intelligence« roots in the Latin verb *intellegere* whose meaning is to gain insight in or cognition of something. This exactly is what computers are *not* able to do; instead, the attribute actually applies to the programmers designing the algorithms such that they fit the computer with intended adaptive behaviour. AI attributed to the system actually is the designers' objectified intelligence, their »coagulated« knowledge and experience, not the system's own achievement. These metaphors appear as linguistic tricks leading people to believe that computers as semiotic machines behave as if they were like humans. In the present discourse on »singularity«, both, euphoric propagandists as well as apocalyptic alerters of unbounded AI, are equally taken in by the self-deception.

This mystification of computers denies the fundamental differences stated by reducing, in a functionalist perspective, competent autonomous acting of humans to algorithmically controlled behaviour of machines. At the same time, it produces illusions about the actual performance capacity of computers. Even worse: The confusion fades out the real problems of the complex relationship between humans and semiotic machines, how human-computer interaction can be made more effective and productive, how computers can be designed as »things that make us

smart« (Norman 1993). How can human action competence and working capacity be enhanced by appropriately designing computer functions such that they meet human action requirements, on the one hand, and how can they be put to effective use in the context of a social practice on the other? The real problem is not to imitate human capabilities, but rather to support and amplify those capabilities by combining them with the performance of computers. This endeavour of human-centred design requires to take needs and conditions of human acting into account and to design computer functions appropriately.

Exactly with respect to this primary and urgent development task in human-computer interaction, present AI efforts generate severe problems rather than enabling solutions. As the behaviour of adaptive systems like multi-agent systems or artificial »neural networks« is history-dependent and based on implicit adaptation processes, it is intransparent and the results are difficult if not impossible to assess. How should humans be able to appropriate such systems with intransparent behaviour, how should they deliberately interact with them, if they behave differently and unexpectedly in comparable situations? Such behaviour contradicts one of the basic rules of human-machine interaction, the requirement of expectation conformity. Without being able to actually assess the validity of the outcomes, humans are ultimately condemned to blindly trust in the system's error-prone performance susceptible to interference – an ethically unacceptable situation (Brödner 2017).

With respect to computer-assisted knowledge work, for instance, as it recently has been envisaged by »cognitive computing« services (by means of IBM's Watson, cf. Kelly 2015; just another misnomer, by the way), these considerations are particularly relevant. Computers in fact appear to be pertinent to collect, store, manage, or retrieve huge amounts of explicit knowledge in various domains. For practical use in specific situations, the knowledge needs to be selected and presented conveniently, a task which typically requires experience and skills in situated judgment power. In situation-specific knowledge application, users assisted by the system are either compelled to blindly bank on its automatic, however intransparent adaptation capacity – as it cannot, for lack of reflecting its own procedures, »explain« or »justify« the validity of its results –, or they need sophisticated interactive software methods and techniques on demand for making the system's calculations transparent on different levels of detail such that they can conclude wether the results are justified (not provided so far).

More generally, the »imitation game« of the »Turing test« is another illustrative example to demonstrate how the functionalist view reduces the rich experience-based being of humans to the behaviour of machines. It is designed to investigate, whether »intelligence« can be justifiably ascribed to a computer as a property of its behaviour. It is based on the exchange of written natural language signs between a person C acting as an interrogator of two players A and B behind a wall where A acts as a man and B as a woman. Player A may be replaced by a

computer for trial. By asking questions of player A and player B, player C tries to determine which of the two is the man and which is the woman. Player A's role is to trick the interrogator into making the wrong decision, while player B attempts to assist the interrogator in making the right one. If the computer succeeds in his role in a sufficient number of cases, it has passed the test (Turing 1950).

Rather than testing artificial »intelligence«, the test actually expresses which relationship human thinking takes to itself: AI conceptions construe the behaviour of objects as if they were intelligent. The Turing test is a form of observing and interpreting sign processes, rather than taking part in a real talk situation: Through the exchange of written signs excluding bodily experience as an essential basis of human perception and intelligence, not really living persons communicate, but rather conventional role models or stereotypes (being a prerequisite for semiotic machines to formally interpret signs). In this arrangement, human experience-based interaction comes down to regular conventional communication patterns, and human intelligence as an individual capability is reduced to a mere property of objects – in other words: Humans in this way construe themselves as machines.

This can be further made clear by looking at theatre performances as a contrasting arrangement where the whole course of events unites actors and audience with their full bodily existence and complete live experience in a common setting. Not just the spoken words taken as signs, but rather the whole situated dramatic procedures and occurrences are implicating the audience and intensifying the actors' performance as well. Based on their different experiences and empathy, some members of the audience may be empathically touched by dramatic situations, others may rather reflect on hidden messages of the drama. All this happens although – or even because – everybody is aware that it is an »imitation game«, a play, not a real drama. It thus perfectly demonstrates what human intelligence is all about.

7 Conclusion: Avoiding the AI Trap

A few important conclusions can finally be drawn from these considerations. First of all, the basic mistake made by the protagonists of the representational world view, from Descartes' *cogito* to present cognitivist and AI communities, consists in regarding conceptual cognition as exclusive access to the world for humans. With this stance, they ignore or even deny the biological roots of human cognition, the existential fact of bodily being-in-the-world with its immediate intuitive perception in situated action.

Prior to conceptual cognition, successful continued action in and interaction with the surrounding physical and socio-cultural world, based on collective intentionality, holistic perception and immediate experience, are at the core of sense-making and human intelligence. It expresses itself as tacit »knowing how« and skillful acting which allocates pre-representational meaning to things, to dealing

with them, and to interactions with others, thus constituting a social practice: Meaning is use (Putnam 1988, Wittgenstein 2009). Explicit propositional knowledge envisaged by cognitivism, in contrast, is derivative only, attained by observing the practice and conceptualising its experiences, it is, therefore, secondary and limited. Due to decontextualisation and abstraction, it has more general validity, but it needs to be appropriated and internalised for practically effective use in situated action, however.

Computer systems, even those being enabled to adapt to environmental conditions by sensor data, attain their functionality solely through conveniently designed algorithms from outside, based on propositional knowledge about their field of application. They, therefore, are lacking own intentionality and self-determined activity as indispensable material basis for perception, sense-making, and experience. With respect to sign processes, they solely operate signals which are, as *quasi*-signs, lacking the references to experienced objects of the world and, hence, cannot »know« for what the signals stand or what they are about. And with respect to abductive reasoning, they are lacking the human capability to create an appropriate hypothesis for transcending the bonds of an existing formal symbol system.

More recent efforts on the part of AI and robotics for »embodying« their systems in order to broaden the range of automatic behaviour do not really change the picture. They all amount to the implementation of purely physical, mainly mechanical or electrical devices enabling sensor-controlled automatic movement in a physical space. This reductionist view of »embodiment« does not with any respect transcend the border to a living body deliberately and autonomously acting in the physical and socio-cultural world around it from which meaning arises.

For these reasons, it is extremely misleading to assign the attribute »artificial intelligence« to computer systems. It distracts the awareness from the fact that all intelligence is on the side of the programmer who provides the computer system with functions that, as its objectifications, solely mimic or simulate intelligent behaviour in a generally limited way. Rather than investing high, but questionable efforts and resources to explore up to which limits such an endeavour can be driven, it appears much more reasonable to thoroughly investigate how the unique human productive forces with its resulting practical and cognitive skills can be enhanced or amplified by combining them with the data processing performance of computer systems.

8 References

Bateson, G. (1980): Mind and Nature. A Necessary Unity, Toronto: Bantam Books

Brödner, P. (2009): The misery of digital organisations and the semiotic nature of IT, AI & Soc 23: 331–351

Brödner, P. (2017): »Super-intelligent« machine: Technological exuberance or the road to subjection, AI & Society Journal of Knowledge, Culture an Communication 2017 (Online first)

Damasio, A. R. (1994): Descartes' Error. Emotion, Reason and the Human Brain, NewYork: G. P. Putnam's Son

Dreyfus, H. (2002): Intelligence without representation—Merleau–Ponty's critique of mental representation the relevance of phenomenology to scientific explanation. Phenomenol Cogn Sci 1: 367–383

Dreyfus, H. L. & Dreyfus, S. E. (1986): Mind over Machine, New York: The Free Press

Dreyfus, H. L. & Taylor, C. (2015): Retrieving Realism, Cambridge (MA): Harvard University Press

Engeström, Y.; Miettinen, R. & Punamäki, R.-L. (eds.) (1999): Perspectives on Activity Theory, New York: Cambridge University Press

Giddens, A. (1984): The Constitution of Society. Outline of the Theory of Structuration, Cambridge: Polity Press

Göranzon, B. & Josefson, I. (eds.) (1988): Knowledge, Skill and Artificial Intelligence, London: Springer

Heidegger, M. (1962): Being and Time, New York: Harper & Row

Husserl, E. (1970): The Crisis of European Sciences and Transcendental Phenomenology. An Introduction to Phenomenological Philosophy, transl. by David Carr, Evanston (IL): Northwestern University Press

Kelly, J. E. (2015): Computing, cognition and the future of knowing. How humans and machines are forging a new age of understanding, Somers (NY): IBM

Leontiev, A. N. (1978): Activity, Consciousness and Personality Englewood Cliffs (NJ): Prentice Hall

Maturana, H. R. & Varela, F. J. (1992): The Tree of Knowledge. The Biological Roots of Human Understanding, Boston (MA): Shambhala Publishing

Merleau-Ponty, M. (1962): Phenomenology of Perception, New York London: Routledge

Nake, F. & Grabowski, S. (2001): Human-Computer Interaction Viewed as Pseudo-Communication, Knowledge-Based Systems 14, 441-447

Nonaka, I. (1994): A Dynamic Theory of Organizational Knowledge Creation, Organization Science 5 (1), 14-37

Norman, D. A. (1993): Things that Make us Smart: Defending Human Attributes in the Age of the Machine, Reading (MA): Addison-Wesley

Nöth, W. (2002): Semiotic Machines, Cybernetics and Human Knowing 9 (1), 5-22

Peirce, C. S. (1903): A Syllabus of Certain Topics of Logic, Boston: Alfred Mudge & Son

Peirce, C. S. (1935): Lectures on Pragmatism, in C. Hartshorne & P. Weiss (eds.): Collected Papers of Charles Sanders Peirce CP 5.141 – 5.212, Cambridge (MA): Harvard University Press

Polanyi, M. (1966): The Tacit Dimension, Garden City (NY): Doubleday

Putnam, H. (1988): Representation and Reality, Cambridge (MA): MIT Press

Pylyshyn, Z. (1984): Computation and Cognition. Towards a Foundation of Cognitive Science, Cambridge (MA): MIT Press

Reckwitz, A. (2002): Toward a Theory of Social Practices: A Development in Culturalist Theorizing, European Journal of Social Theory 5 (2), 243-263

Rohde, M.; Brödner, P.; Stevens, G.; Betz, M. & Wulf, V. (2016): Grounded Design – a Praxeological IS Research Perspective, JIT 2016.5 AOP

Ryle, G .(1949): The Concept of Mind, London: Hutchinson

Schon, D. A. (1983): The Reflective Practitioner: How Professionals Think in Action, New York: Basic Books

Searle, J. R. (2010): Making the Social World. The Structure of Human Civilization, Oxford: Oxford University Press

Suchman, L. (1987): Plans and Situated Actions. The Problem of Human Machine Communication, Cambridge (MA): Cambridge University Press

Tomasello, M. (2014): A Natural History of Human Thinking, Cambridge (MA): Harvard University Press

Tomasello, M. (2008): Origins of Human Communication, Cambridge (MA): MIT Press

Turing, A. (1950): Computing machinery and intelligence. Oxford University Press, Oxford

Varela, F. J.; Thompson, E. & Rosch, E. (1991): The Embodied Mind. Cognitive Science and Human Experience. Cambridge (MA): MIT Press

Winograd, T. & Flores, F. (1986): Understanding Computers and Cognition. A New Foundation for Design, Norwood: Ablex Publ.

Wittgenstein, L. (2009): Philosohical Investigations, Chichester: Wiley-Blackwell

↑Inhalt↑

»Super-intelligent« Machine: Technological Exuberance or the Road to Subjection

Abstract: Looking back on the development of computer technology, particularly in the context of manufacturing, we can distinguish three big waves of technological exuberance with a wave length of roughly 30 years: In the first wave, during the 1950es, mainframe computers at that time were conceptualized as »electronic brains« and envisaged as central control unit of an »automatic factory« (Wiener). Thirty years later, during the 1980es, knowledge-based systems in computer-integrated manufacturing (CIM) were adored as the computational core of the »unmanned factory«. Both waves dismally stranded on the contumacies of reality. Nevertheless, again thirty years later, we now experience the departure of the »smart factory« based on networks of »artificially intelligent« multi-agent or »cyber-physical systems« (often also addressed as »internet of things«). From the very beginning, these technological exuberances rooted in mistaken metaphors describing the artifacts (e.g. »electronic brain«, »knowledge-based« or »intelligent systems«) and, hence, in delusions about the true nature of computer systems. The behaviour of computers is, as computing science teaches us, strictly restrained to executing computable functions by means of algorithms, it thus neither resembles the performance of a brain as part of a complex sensitive living body nor is it in any meaningful sense »knowledgeable« or »intelligent« (this predicate remaining reserved for the programmer designing the algorithms). When the delusion of being able to implement »smart factories«, despite the countless accomplishment failures before, gains momentum anew, it appears absolutely essential to reflect on underlying misconceptions.

Key words: Automatic factory, cyber-physical systems, multi-agent systems, artificial neural networks, big data, functionalism, praxeological perspective.

1 Introduction: Dreaming of the Automatic Factory

With the ubiquitous proclamation of »industry 4.0« and »big data« as core of a new »industrial revolution«, we experience another wave of technological exuberance propagating advanced computer technology as a panacea for all sorts of societal problems from poor resource efficiency to demographic imbalances. Hardly any trouble appears big enough not to be overcome by »digitization«. Against the background of knowledge-intensive value creation, continuously weak growth and decreasing productivity growth, the World Economic Forum reminds in a report, to undertake major efforts specifically for gaining higher

competitiveness in manufacturing and services by improving performance and control over globally dispersed value adding chains by means of the »digital transformation « – seemingly the original document of the movement (WEF 2012).

In accord with these widely shared ideas, highly developed and industrialized countries adopted more or less effective measures for their advancement; particularly, Germany with its strong industry base launched a framework programme for the advancement of »innovation in manufacturing and services« as central part of their »high tech« development strategy. Production components such as »intelligent« machines, work pieces, and storage systems are envisaged to form globally networked »multi-agent« or »cyber-physical systems« (CPS). Enabled by advanced computer technology, these systems can automatically exchange data and mutually initiate actions between the components in »decentralized self-organisation« and, thus, accomplish »smart factories and services«, sufficiently adaptive and dynamic for economically producing individual customer orders, handling disturbances and failures, and optimal decision making (BMBF 2014).

More generally, in their new book Brynjolfsson and McAfee (2014) describe the advent of a »second machine age« and how the new digital revolution changes the world. Referring to the extraordinary exponential growth of computer performance and storage capacity in digital networks (according to Moore's law), with rapid progress in accomplishing »artificial intelligence« and »big data« applications, they illuminate the potential for »digital« value creation. For explanation, they refer to Google's the self-driving car and IBM's Watson with respect to knowledge processing capacities. In critical perspective, they also address, however, the risks of monopolizing digital value creation due to network effects (»the winner takes it all«), look at expanding inequalities and polarities of skills and income, and discuss opportunities for controlling future developments.

This in many respects evokes memories of earlier attempts to make come true management's old dream of an automatic factory, of eternal unmanned »value creation« by means of computer and sensor technology, the dream of finally becoming independent of the obstinacy and contumacy of living labour. As early as 1950, when the first commercial mainframe computers had just been installed, Norbert Wiener (1950) already had a clear and detailed vision how to achieve an automatic factory by means of sensors, effectors and computing machines as central logical units for controlling its complex processes. And thirty years later, during the 1980es, the central idea of the »unmanned factory« directed high development efforts into »knowledge-based« (i.e. equipped with symbolic »artificial intelligence«) and computer-integrated manufacturing (CIM) systems (Brödner 1990, Hunt 1989). These tidal waves of technological exuberance, arriving with a length of roughly 30 years, each time dismally stranded at the cliffs of unruly matter and underestimated implementation problems with ensuing long phases of disillusionment when trying to overcome those difficulties. Ironically, in each of

these sobering phases, the value of implicit knowledge, of intuition and creativity, of specifically human acting skills was rediscovered.

The paper wants to realistically assess the similarities and differences of the new promises of the »second machine age«, particularly »industry 4.0« and »big data«, relative to previous attempts. To this end, the paper starts with presenting the scientific and technical foundations of the various attempts to accomplish an automatic factory until present in some more detail. By comparing the attempts, the novelty of the most recent approach can be determined. Based on that, a critical evaluation of the opportunity and risk potentials can be made. Finally, realistic design perspectives for forward looking manufacturing and service systems will be derived.

2 Revenant Symptoms: The Third Wave

2.1 Previous attempts of creating an »unmanned factory«

Looking back on the development of computer technology, particularly in the context of manufacturing, we can distinguish three big waves of technological exuberance with a wave length of roughly 30 years: In the first wave, during the 1950es, mainframe computers at that time were conceptualized as »electronic brains« and envisaged as central control unit of an »automatic factory«:

> *»The computing machine represents the center of the factory, but it will never be the whole factory. On one hand, it receives its detailed instructions from elements of the nature of sense organs. […] Besides these sense organs, the control system must contain effectors or components which act on the outer world. Of course, we assume that the instruments which act as sense organs record not only the original state of the work, but also the result of all previous processes. Thus the machine may carry out feedback operations, either those of the simple type now so thoroughly understood, or those involving more complicated processes of discrimination, regulated by the central control as a logical or mathematical system. In other words, the all-over system will correspond to the complete animal with sense organs, effectors, and proprioceptors, and not, as in the ultra-rapid computing machine, to an isolated brain, dependent for its experiences and for its effectiveness on our intervention«* (Wiener 1950, 156f).

Despite the persuasive power of Wiener's clearly outlined vision of the automatic factory, computers did not penetrate manufacturing to a considerable degree until the mid 1970es.

There were of course, some cases of early investment in, for instance, numerically controlled machine tools and computerized management of a firm's material (and money) flows as well as R&D activities in computer-aided design (all being areas with already highly standardized operations and procedures). However, difficulties in getting access to isolated mainframe computers prevented wide-

spread use. With the advent of »virtual machine « operating systems (IBM 360/370, DEC PDP 10/11) providing computer access via locally dispersed terminals, the use of computing power in manufacturing gained considerable momentum. Far from implementing fully automated operations , however, computing machinery was rather used in a more or less interactive mode combining computing functions with skilled human expert work.

Meanwhile, during the high time of Taylorism with its separation of planning from operating, huge amounts of explicit propositional knowledge about optimal operating conditions and procedures in manufacturing had been collected. Additionally, it turned out that, with rapidly expanding computer programs in many manufacturing areas (NC machines, computer-aided production planning and control, computer-aided design, cost accounting), many of these programs used the same data. In order to avoid error-prone multiple data entries, the idea arose to integrate the many programs and software components so far used in isolation into »computer-integrated manufacturing« systems (CIM) by means of a common data base. Moreover, as at the same time market forces and competition changed their nature from standardized mass to flexible quality production, and against the background of a wealth of explicit manufacturing knowledge at hand, the integration idea was combined with efforts to develop symbolic artificial intelligence using this knowledge for automatically handling the complex and dynamic, steadily changing operating procedures. The intention was to widely replace skilled shop-floor and knowledge workers by »knowledge-based systems«. »Experts leave, while expert systems remain« was a common slogan at the time. This is how thirty years after Wiener's vision, during the 1980es, »knowledge-based« and »expert systems« in computer-integrated manufacturing were promoted as guiding ideas and computational core of the newly envisaged »unmanned factory« (for more details cf. Brödner 1990, 2007, Hunt 1989).

Both waves were fueled by a technology-centred perspective which culpably ignored essential conditions for successful performance of manufacturing processes. In particular, it disregarded deep societal changes such as the transition from industrial to knowledge-based economies and, hence, the relevance of social relationships and the division of labour and knowledge for efficient value creation. As knowledge work is becoming a dominant factor in manufacturing and services (Bell 1973, Drucker 1994), it is important to get a clear understanding of the fundamental differences between implicit practical competence and explicit conceptual or propositional knowledge and how they interplay with each other (for basic differences cf. Nonaka 1996, Polanyi 1966, Ryle 1949). While the individually embodied action competence as pre-reflexive implicit knowledge or working capacity is always antecedent and expresses itself in activities of successful social practice, explicit propositional knowledge *about* certain aspects of this practice can only be gained through observation, concept formation and analysis. However, this codified knowledge needs to be made effective again by appropriat-

ing it for practical use (this also holds for technical artifacts derived from this knowledge). Both, explicating propositional knowledge about competences and appropriating it for practical use, are skillful activities expressing the working capacity which for his part is augmented through these activities. Ironically and contrary to common expectations, these dynamic relationships have just precisely been illuminated by the difficulties of »knowledge elicitation« for building expert systems (Brödner 2013).

Consequently, both waves dismally stranded on the contumacies of reality. In fact, the efforts to build knowledge-based integrated manufacturing systems ended in a complete reversal: Confronted with examples of real high performance manufacturing systems, it became obvious that these systems, contrary to common belief, produced their high efficiency not primarily by means of advanced computer technology, but rather through skillful cooperation between human experts in multifunctional teams such as cellular group work or simultaneous engineering teams. Exactly this could also be concluded from theoretical insight in the dynamic relationship between the – always partial – explication of practical experience and action competence into codified knowledge and the appropriation of this knowledge for practical problem solving. The latter typically requires expertise and knowledge from diverse domains to be consolidated and integrated through the self-organized cooperation of experts. The more differentiated, complex, and dynamic the codified knowledge is – and its objectification in technical artifacts –, the more demanding competence and working capacity are required to seize hold of these productive forces for effective practical use. This is subject to the experts' autonomy and cannot be planned and instructed.

Both waves were also accompanied by apocalyptical predictions of lasting technological unemployment. As the shop-floor worker before, now also the knowledge worker would be replaced to a substantial degree by computing machinery. Electronic data processing was generally denounced as severe »job killer«. Nothing of this really happened, though; instead, a productivity paradox could be observed with computer use: »You can see the the computer age everywhere, but in productivity statistics« (Solow 1987). Although the rapidly expanding use of computers in manufacturing and services, growing both in volume and diversity of applications, caused massive changes in professions, specific skills, and qualifications, the macroeconomic productivity effect was minimal. In fact, we experience, since a number of years, a secular downturn in macroeconomic productivity growth rates (Gordon 2014). On firm level, however, huge differences in performance between firms operating under comparable market conditions and with similar software applications in use could be observed. Two decades of empirical research efforts investigating these effects finally came, in accord with theoretical view above, to the conclusion: »To leverage information technology investments successfully, firms must typically make large complementary investments and innovations in areas such as business organization,

workplace practices, human capital, and intangible capital.« (Jorgenson et al. 2008, 10; similarly also Dedrick et al. 2003).

2.2 The new machinists' claims

Notwithstanding these rather sobering experiences, we now see, again thirty years later, the departure of a »smart factory« based on networks of »artificially intelligent« multi-agent or »cyber-physical systems« (CPS; often also addressed as »internet of things«). With the striking designations »industry 4.0« or »second machine age«, respectively, this is intended to mark another qualitative leap in industrial development. With its focus on advances in computer technology, it again indicates a new wave of techno-centrism and technological exuberance. So it is worth while to throw a closer look on the scientific and technological foundations.

Embedded systems are computer components for digital control of physical processes which are equipped with interfaces to humans and other components. By data exchange via the internet, they can be globally networked (cyber-physical systems«, »internet of things and services« (cf. fig. 1).

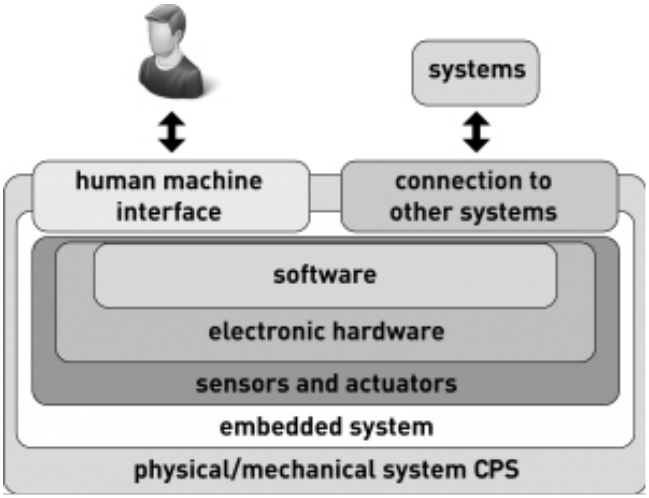


Fig. 1: *Embedded system (according to Broy)*

Multi-agent systems (MAS, also called »distributed artificial intelligence«) consist of software- agents with limited autonomy for goal-oriented interaction by data transfer; such concerted action enables them to jointly master demanding tasks.

For a deeper understanding, a closer look on the behaviour of these systems, their mode of operation and the way they are described is required. The etraordinary interest in MAS goes back to the idea that coordinated action of a large number of units with relative simple behaviour would produce »artificial intelligence« (Minsky 1988). It is based on the explicitly articulated conviction that »interaction is more powerful than algorithms« (Wegner 1997). This has, however, immediately been proven wrong, as MAS also underly the constraints of comput-

ability (Prasse & Rittgen 1998). As a matter of fact, however, the behaviour of MAS as wholes does show emergent properties that cannot be observed with any single software-agent.

Agents are software engineering objects capable of taking in sensor data from the environment as well as from other agents, of independently processing the data by means of their own algorithms – mostly »machine learning« algorithms –, and of putting out resulting data. The behaviour is characterized by the capacity to follow goals and to adapt to changing conditions by »machine learning« (Breadshaw 1997, Maes 1994). In order to cope with more demanding tasks, the agents with limited autonomy each can thus cooperate for achieving the tasks by concerted action (Wooldridge 2002). Simultaneously, huge amounts of data are being generated which can be used separately.

Although each agent by itself performs relatively simple algorithms and shows transparent behaviour, the MAS as a whole owns a highly complex behaviour which cannot be analysed and understood from outside any more, although it still is strictly determined by algorithms. Formally, CPS and MAS can be described as so called »non-trivial machines« (Foerster 1991) whose output data are not only determined by its input data, but also by its variable internal state that is itself a function of input data. The internal state reflects the various ways in which the agents interact and adapt their behaviour (cf. fig. 2). Consequently, the MAS behaviour highly depends on the history and cannot be analytically determined from outside and, hence, foreseen any more.

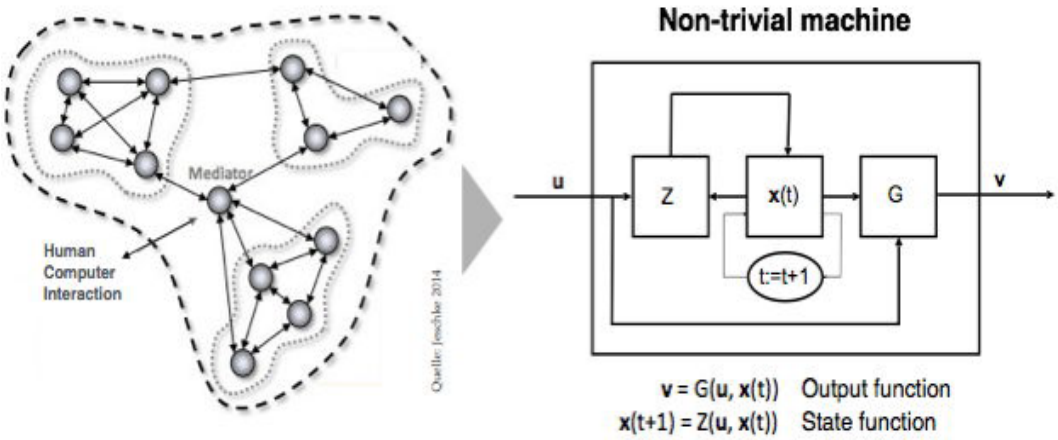


Fig. 2: *MAS as non-trivial machine*

Besides this physical and algorithmic description of the MAS behaviour, another type of description is also commonly used which orients itself at the agents' purposive interactions according to the so called »intentional stance« (Dennett 1987). In this stance, intentional states like convictions, beliefs, desires or intentions are ascribed to agents expressively in order to simplify the description of the agents' behaviour. This common practice among MAS researchers refers to a report by McCarthy (1979):

> *»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. Theories of belief, knowledge and wanting can be constructed for machines in a simpler setting than for humans, and later applied to humans. Ascription of mental qualities is most straightforward for machines of known structure such as thermostats and computer operating systems, but is most useful when applied to entities whose structure is incompletely known.«* (McCarthy 1979, quoted according to Shoham 1993, 53).

An agent's intentions are formally described by means of propositional modal logic which augments the normal two-value logic (a proposition can be either true or false) by the two modalities that a proposition is necessarily or only possibly true or false. On this augmented logical basis, an agent's intentional states like beliefs, desires or intentions can then be expressed in a formal language. The fact e.g. that agent a has the conviction that proposition ϕ is true at time t can thus be modelled by the formal language expression $Bel(a,\phi)(t)$ (Wooldridge 2002).

The abstraction from real physical and algorithmic behaviour by ascribing intentionality to software agents or machines is an attempt to escape from the unpleasant fact that the course of this behaviour is intransparent, although determined. Because the behaviour cannot be explained on the physical level, it is pretended, by abstraction, to underly intentionality.

This lastly absurd kind of quasi-explanation leads to believe only that comprehension is possible, while the real behaviour still defies understanding. Rather than analyzing the problem, it is obscured instead. Both, however, the missing behavioural transparency as well as the attempt of ascribing intentionality as quasi-explanation, have fatal consequences for human-machine interaction with and safety of MAS (Norman 1994).

The vision of the »smart factory« highly depends on »machine learning« capabilities for adaptive behavior. They comprise diverse methods and algorithmic procedures for purposefully changing an agent's structure, or its program respectively, such that its behavior is improved relative to a given utility function. The present chief attraction of the »smart factory « and »smart service« propagandists is »deep learning« in so-called »artificial neural networks« (ANN; for an introduction cf. Kriesel 2007). They consist of many simple, connected processors called neurons, each producing a sequence of real-valued activations according to specific computing functions resembling neural functions (cf. fig. 3).

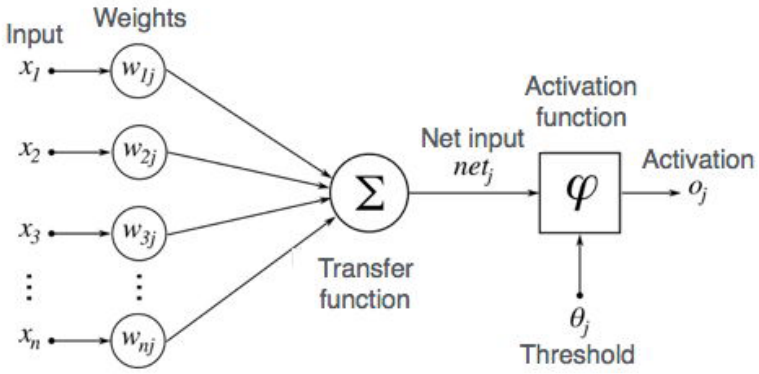


Fig. 3: *Computing functions of a «neural« processor forming a node*

Input neurons get activated through sensors perceiving the environment, while other neurons get activated in a layered order through weighted connections from previously active neurons. Some neurons may influence the environment by triggering actions. »Learning« then is about finding weights that make the neural network exhibit desired behaviour.

Depending on the problem and how the neurons are connected, such behaviour may require long sequences of algorithmically controlled computational stages, where each stage transforms the aggregate activation of the network. For problems of speech recognition or image categorization e.g. – forming tasks where ANN are specifically successful – a very long sequence of patterns or images is presented as input together with the correct categories at the output; from this assignment, the network can, in many small adaptive steps, automatically compute the connecting weights w_{ij} (which implicitly mirror the learning progress) (cf. fig. 4).

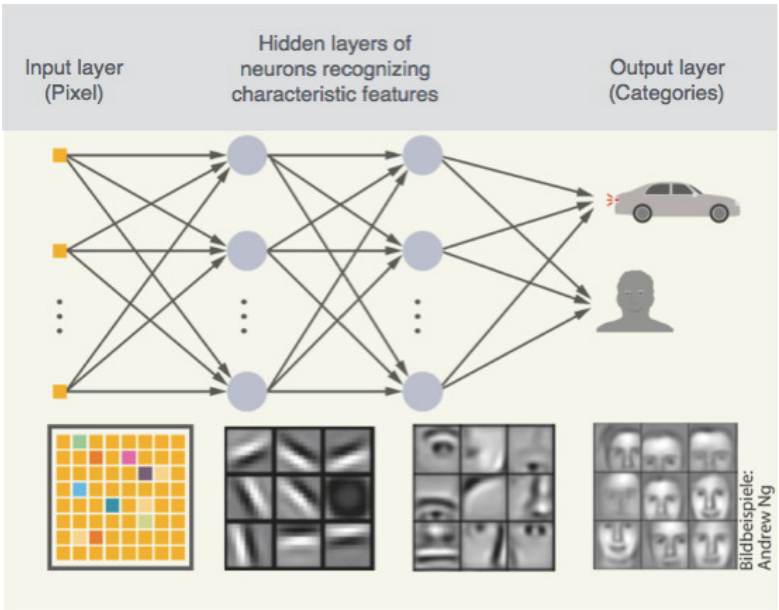


Fig. 4: *ANN for categorizing images (Source: c't 6/2016)*

Consequently, successful ANN deployment does not depend on analytical insight into cause and effect relationships but rather is the result of theory-less trial and error with different structures and learning algorithms for adjusting the weights. Due to their extraordinary performance, so called »convolutional neural networks« whose structures are inspired by biological models have moved into the focus of interest for the time being.

Contrary to what is suggested, ANN are not conceptually new computational devices at all – in fact, they first appeared in 1950es and have been developed and proven since (with a long »winter« of disinterest in-between). Progress is not accomplished by any conceptually new AI idea but predominantly by the exponentially increased computational power allowing for much larger networks with considerably more components and layers (obviously in the illusion that quantity somehow turns into quality). Ironically, good results are mainly achieved, however, by the intuition, skills, and experience of the developers in structuring the networks and mastering the numerous practical computational problems – from the »vanishing gradient« to the variable increment control of learning algorithms (Schmidhuber 2015).

In practical operation, users have to cope with the uncertainty whether the computed results are correct and suitable in the long run. Even if the ANN deliver adequate results in the vast majority of cases, they may suddenly fail without notice by the users. Even slightly disturbed input data can lead to considerable failures (Sharif et al. 2016). For lack of transparency of the non-linear behavior of ANN, its reliability is difficult to evaluate; basically, users have no chance but to blindly trust in their functionality.

Moreover, the appropriation for interactive use is seriously troubled in case of multi-agent systems with »deep learning« facility. Formally, these systems are »non-trivial machines« whose behaviour depends on history and, therefore, is intransparent and unforeseeable, although algorithmically determined. How can human actors put such systems to deliberate instrumental use that each time exhibit a different behaviour – a property that clearly contradicts the HCI requirement of expectation conformity? On the users' side, excessive expectations for the systems' alleged »action competence« would be evoked at the same time. Confronted with this kind of contradictions, simultaneously exposed to high pressure of management's expectations for successfully mastering their tasks, despite the loss of control over their means of work with intransparent behaviour, the workers would suffer from permanent psychic stress (as already analyzed by Norman 1994).

Moreover, the division of tasks and functions between remaining workers and automatically operating computing components is of fundamental significance for sociotechnical systems design. With respect to this essential design aspect, a number of »ironies of automation « have been early revealed from analyzing working activities in central control rooms, which even gain relevance since with growing

systems complexity: Automatically operating »learning« systems like MAS are designed to widely replace human expert knowledge workers, whose working capacity is, however, urgently needed in cases of disturbance or failure. The specific skills of the working capacity are fading away, though, to the extent they are not used during normal automatic operations. In the long run, a severe loss of competence will occur which will turn originally highly competent users into helpless, unpracticed »operators« (Bainbridge 1983, for impressive new examples cf. Carr 2013).

Consequently, these systems are inappropriate for interactive use; they can only be designed and operated as self-contained automata with the incalculable risk of undesired behaviour (examples of such »normal accidents« (Perrow 1984) are numerous).

2.3 Persistent self-deception

In order to catch the original speech flair of the »smart factory« propagandists, it is worth while to start with quoting a typical representative, a top manager from a German automotive company, who at a recent conference characterized the specific features of the »Robotic Enterprise: The Future AI Company« in the year 2025 with the following statements:

> *»Super intelligent, continuously learning computers will take over much of what humans deal with so far: They automatically respond to customer or supplier questions by means of so called bots, they autonomously decide how prices for more than 200 models are adapted from country to country, they even design cars and compute how they can be produced. They will also take over cost accounting and controlling. […] They will even manage business meetings«* (SZ 11.11.2016, own translation).

Comparing this with a similar statement from the previous wave of technological exuberance (or again with Wiener's earlier vision of 1950 quoted above) reveals high accordance:

> *»Computer integration represents the core of the future manufacturing innovation. It aims at automatically producing variable production programmes. … A new manufacturing structure emerges which, as a mechanic organism with programmed and, hence, stored intelligence is capable of automatically producing goods. … On this higher development level, the factory will need machine intelligence«* (G. Spur 1984, a leading German manufacturing researcher at the time; own translation).

An unmistakable red thread of obviously contrafactual and wishful thinking winds through these visions. From the very beginning, these illusions and technological exuberances rooted in mistaken metaphors describing the envisaged computer artifacts, e.g. »electronic brain«, »artificial intelligence«, »knowledge-based systems«, or »machine learning«. With their analogies and references to specifically human capabilities, the metaphors obscure essential differences between artificially created machines and autonomously living, socially interactive organisms.

Hence, they produce delusions about the true nature of computer systems. The behaviour of computers is, as computing science teaches us, strictly restrained to executing computable functions by means of algorithms, it thus neither resembles the performance of a brain as part of a complex sensitive living body nor is it in any meaningful sense »knowledgeable« or »intelligent« – this predicate remaining reserved for the programmer designing the algorithms or the users making sense of the computing functions. When the delusion of being able to implement »smart factories«, despite the countless accomplishment failures before, gains momentum anew, it appears absolutely essential to reflect on underlying misconceptions. According to Hofstadter & Sander (2013), analogies are at the core of cognition; analogies allow to understand encountered new phenomena by means of existing experiences, they are instruments by which we apply the wealth of our previous experiences to the presence, and without them we would helplessly navigate in the world. Therefore, it is of essential importance to draw on appropriate analogies transferring the predominant characteristic to the new phenomenon. Exactly this fails with the above analogies taking specific human capacities for machine functions, thus confusing the true nature of both. This fallacy ultimately leads to a mistaken equating of both phenomena.

This can be exemplified by ascribing intentionality to machines according to the »intentional stance« (as quoted above). It is legitimated »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.« The key word here is »information« which itself is totally confusing, as it denominates different, incompatible concepts: either the syntactical measure of the »entropy« of a string of signs from a finite set (alphabet) according to Shannon (1948) or »any difference that makes a difference« in the context of a social practice according to Bateson (1980). By leaving this open, the physical world of deliberately designed machines with prescribed behaviour is confused with the social world of autonomous actors with the faculty of speech, of creating knowledge, and of designing purposeful artifacts.

Similarly, the term »machine learning« is again based on a mistaken analogy or attribution. The machine's changing behaviour is achieved by algorithmic procedures controlling its adaptation to environmental changes (in fact, this type of machines have formerly been rightly called »adaptive systems«). In contrast, human learning is essentially based on reflective action control and the capacity of concept formation as foundation for creating explicit propositional knowledge. This confusion nourishes the illusion that computers equipped with »artificial intelligence« and »deep learning« capacities can widely replace human skills and working capacity.

In philosophical terms, these mistaken metaphors are built on the fluctuant grounds of *functionalism*. Elaborated as an approach to overcome the flaws of be-

haviourism, functionalism (Putnam 1960, Fodor 1968) recognises mental states as essential internal entities for explaining behaviour. Irrespective of their material implementation (as electronic hardware or a biological brain), the mental states are regarded, however, as purely functional states according to the Turing machine model. This philosophical view is meanwhile resumed to be refuted, though, because equal mental functions can – as Putnam, one of its originators later has shown (1991) with sharp-witted thought experiments – produce totally different real world references like thoughts or experiences. In its more recent variation, with an interpretation of »embodiment« (cf. e.g. Varela et al. 1991) that again is arrowed by a positivistic attitude, CPS are simply equated with sensitive living human bodies which are up to empathy and able to reflect their context-bound experiences (cf. e.g. most recently Jeschke 2015). The scientifically founded difference between algorithmically determined behaviour of an artifact and intentionally controlled sense making action and understanding in the context of social practices is again ignored.

With respect to the use of computers in organizations with knowledge work or value creation, where human experts interact with advanced computer systems, this difference is highly relevant for assessing the machinists' claims of being able to implement »smart factories or services« replacing human competence and expertise. At the core, they deny the fundamental ontological difference between physical events and social facts: While causal relationships in the physical world – in which, on the basis of semiconductor physics and formal logic, machine computation is operating – exist independently of human activity, objects and facts of the social world such as signification, meaning, or institutions are solely created and maintained through communication and cooperation based on shared collective intentionality: They are originated by declaration, i.e. by speech acts that make something the case just by representing it as being the case (Searle 2010).

Consequently, there seemingly is a unsurmountable gap between both worlds with respect to the social construction of meaning being inaccessible for computing machinery unless it is embedded in and appropriated by actors in the social world of an organization. For bridging the gap, it is useful to pick up the triadic sign concept elaborated by C.S. Peirce (1903), an American logician rooted in the pragmatist strand of research (who also was the first to develop a first order predicate calculus as another foundation for computing science).

With this sign concept, he distinguishes a physical entity, a signifying element called representamen, from a second entity as the object it refers to or designates, which can be present, distant, or imagined only. A third entity, the interpretant, assigns meaning to this reference in some context of social interaction. This sign concept provides the missing link to connect the physical world, in which computers process signals or data, with the social world of signification, of constructing meaning through interpretation in an action context.

Distinct signification or interpretation of the data are possible due to the actors' functional knowledge of the algorithms and the inputs. Specifically, it opens up a *praxeological perspective* (Reckwitz 2002) for analyzing the complex interplay of algorithmically determined physical data processing with the social process of signifying or interpreting the data in the context of an organization's social practices (this can be done e.g. leaning on Giddens' theory of structuration (1984), for more details cf. Brödner 2009). In this way, the triadic sign concept mediates between signal and sense.

Confusing both worlds – as it is indicated for instance, when meaningless data are constantly equated with meaningful information – has effects in two directions: In one direction, it reduces living beings, in a positivistic and reductionistic attitude, to the functionality of machines, while on the side of the users it creates the illusion of capabilities comparable to theirs. Leaning on this, the »smart factory« approach can, of course, be misused as ideological offensive: With threatening scenarios of replacement, living labour is set under pressure to accept the new »industrial revolution« in all aspects as an inescapable »natural « event. Public awareness thus is distracted from deliberate massive institutional, particularly labour market deregulation which often are the real threat to decent work and income (which is e.g. mostly the case with businesses operating »crowdsourcing« and »clowdworking« platforms). In this way, detrimental consequences can easily be ascribed to »technological progress« which allegedly cannot be »hold up«.

3 Big Data and the Struggle for Autonomy

The pervasive implementation of »cyber-physical systems« is accompanied by origination of huge amounts of data – »big data« – being processed in unprecedented volume, variety, and velocity. These data can originate from diverse sources and they may differ in the way they are structured, e.g. as text or image documents, or data base entries, and can still be combined for processing. Because of the exponential increase in computer power with respect to processor and storage capacity, it is now possible to keep huge amounts of data ready for very fast processing in the random access memory (so called »in-memory technology«). Moreover, highly expanded band widths allow for transferring huge data volumes.

If necessary and with suitable tasks, data volumes and algorithms can even be split to different, locally dispersed processors. All this contributes to exploiting the full performance potential of advanced computer technology for coping with complex tasks.

Big data processing gives rise to some substantial, eventually even unsolvable problems, though. One severe problem concerns the methodology of rational processing itself. Only recently, the editor of the internet magazine »Wired« has, in typical, commonly shared attitude of technological exuberance, proclaimed the »end of theory« in a full-bodied way: Theoretically informed research with

proven scientific methods could just be replaced by huge volumes of data, in the »petabyte age« forecasts on the basis of pure correlations would be superior to hypotheses-based propositions, and correlation would replace causality (Anderson 2008).

This unbelievable folly reproduces the well-known fallacy of *»cum hoc ergo propter hoc«*: When two events *a* and *b* coincide, one can never know without additional expensive analysis whether *a* has been caused by *b* or reversely *b* by *a*; one cannot even know whether both events depend on an unrecognized common third incident or whether they occur just by accident. Ultimately, this view leads to the apophenic delusion of perceiving patterns in purely random data. It looks like the inmates are running the asylum.

Big data volumes, particularly if they stem from different sources, normally have deficient quality: The data mostly are not representative or error-prone, they can even be obsolete or inconsistent. In many cases, one cannot even assess the extent to which the quality is deficient.

As long as the big data processing does not apply accepted strong methods for statistical conclusion, however, which include knowledge of the data quality, it must be seen scientifically embellished reading tea leaves. Nevertheless, data have frequently been adjudged to be an »important economic good«, the »bulk oil of the 21st century«; if so, they then need equal careful and expensive refinery for extracting useful information. Finally the deficient data security produces severe problems. Organizations run into huge risks by loss or theft of data, either spying from outside or sabotaging from inside (risks about which almost daily reports on »cyber attacks give evidence). The risks become even higher, if the data and the processing procedures use to be outsourced to service providers or into the »cloud«. With respect to frequency and volume of the damages experienced, it is hard to understand why firms with ambitions for »industry 4.0« projects deliberately expose their continuously emerging, highly competition-sensitive data streams about products and processes to such risks. Even if much technical and organizational efforts is invested in data security, they will never be sufficient, since any sophisticated security measure can, as experience teaches, be overcome again.

According to a bon mot of the grand semiotician Umberto Eco (1976), a sign is »anything that can be used for lying«. This illuminates in a paradoxical way again the deep insight Peirce had into the logic of signs according to which data represent something in a certain context and for a social actor only. Their meaning always is the result of interpretation in the shared context of a social practice. That which a physical sign stands for, the designated object, and the meaning which is ascribed to it, are first of all up to the sign's author using it for a message. Those who take it up for interpretation are free, to interpret it as expected or other (as far as the context allows). In other words: How data are being interpreted defies the author's control. This is why signs can always be used to deceive, to trick, to de-

fame, or to degrade (all this being frequent practices in the social web and by secret services as well). Via the social web, thus otherwise locally constrained practices of intrigue are becoming a global phenomenon. Physical computer signals or data are, due to their formal and abstract nature, context-free and meaningless; nonetheless, they frequently use to be equated with meaningful information derived from context-dependent interpretation. By this common error it is suggested that physical signals or data as such allegedly possess information, meaning, and validity quasi as fixed qualities. It is true that parts of the context can be reconstructed from a number of data referring to the same object or person if the according algorithm's semantics is known and, thus, constrain the range of interpretation, but this incompletely reconstructed context still leaves space for various other interpretations and misinterpretations.

Despite this interpretive space and often questionable data quality, data suggest objectivity and factuality like in cases of presumptive evidence (rightly seen as questionable). Moreover, due to the social construction of reality, the interpretation often describes a reality just created by the signifying process itself or it even unfolds the effect of a social norm: Descriptive can be turned into normative data, frequency can change to certainty, and interest- bound signification can be enforced by power (as can be seen e.g. in cases of self-tracking or of determining creditworthiness; cf. Boyd 2011).

Due to these peculiarities of the social use of signs, control over the data and their processing algorithms delivers a powerful ruling instrument in the hands of management or government. With comprehensive global data collections and various data processing methods, a powerful instrument for behavioural control and dominance emerges in the hands of the owners; it can, at any time, be used in many ways for exerting influence and power at their discretion, from manipulating public opinion to threat and blackmail. This is possible exactly because the suppliers of the data loose the sovereignty and control over their interpretation in the moment they give the data away. This enables the successful implementation of a perfect panopticon (*sensu* Bentham and Foucault) lighting all corners of knowledge work processes or the social web: It remains the secret of the observer whose behaviour he observes and how he is interpreting it.

Besides the severe use problems with non-trivial machines, this loss of informational autonomy appears as the most threatening societal damage the third wave is about to produce. As it might end in a digital totalitarianism, the struggle for autonomy on all levels of social practice is of foremost importance.

4 Conclusion: Perspectives for Intelligence Amplification

Computers are data processing machines, hence their functionality is semiotic in nature. They fundamentally differ from classical machines transforming matter or energy: While the latter operate in the physical world of natural processes and

their functionality makes use of natural forces and effects for increased efficiency and productivity, computers perform computable functions within formalised sign structures, processing signals or data determined by algorithms, nothing else. Formalizing sign processes, their reduction to computable functions, therefore is a necessary prerequisite. When operating in organisations, computers and their »auto-operational forms« (Floyd 2002) are, based on sufficiently modelling and formalizing underlying sign processes, fully embedded in the social world of social interaction in the organization's practice, namely the expanding knowledge work. Computers can, thus, be used to organize, process, and store codified knowledge represented in data (Brödner 2009). Productivity, therefore, can only improve, if the sign processes of this social practice are organized more efficiently through computer operations – this being the true reason for the empirical findings on the productivity paradox quoted above. Unfortunately, these essential relationships are obscured by the unreasonable terms »digitization« and »digital transformation« permanently used for computerizing knowledge work.

Only by emphasizing the fundamental differences, it is possible to adequately design decent and efficient computer-supported work. Recognizing the differences and taking the praxeological perspective as outlined allows to focus the view on how exactly computer artifacts are emerging from conceptually analyzing social practices, how appropriating their functions for effective use intervenes in social practices of knowledge work, and what the decisive issues of taking influence are. Both, the design, predominantly the analysis, modelling, and formalization of sign processes, as well as the organization of the elaborate appropriation of the computer functions derived from that for effective practical use, are the neuralgic fields of participatory intervention of computer experts and potential users. Both are highly contested terrains with respect to interpretation, interests and exertion of power. That is why they also are the main fields of influence of knowledge workers and their stakeholders. Due to its utmost importance as productive force, the development of working capacity of living labour, its implicit knowledge, practical skills, and competences, must be the guiding principle.

In design of sociotechnical systems, in particular in development, implementation, and use of computers in manufacturing, activities have, as historical retrospective shows, always been underlying two contrarian perspectives:

- The *technology-centred perspective* of most extensively automating knowledge work as it is driven by the efforts for accomplishing »artificial intelligence« – *AI (artificial intelligence)*: »Smart machines« and »autonomous agents« networked as multi-agent systems with »deep learning« capacity and combined with »big data« procedures are envisaged to imitate and widely replace human working capacity in manufacturing and services; their capacity to »learn« – in fact their adaptivity to environmental conditions only – is nevertheless supposed to provide sufficient flexibility for adapting to chan-

ging requirements (according to the »intentional stance«; Minsky 1988, Shoham 1993, Wooldridge 2002).

- Under the *praxeological perspective*, in contrast, advanced computer systems, designed, appropriated, and used as human and task appropriate tools and media for cooperation – *IA (intelligence amplification)* –, are envisaged to support living labour in such a way that the working capacity and, consequently, productive and innovative capacities are enabled and stimulated to grow: »things that make us smart« (Norman 1993; cf. also Ehn 1988, Winograd 1996).

As indicated above, due to poor previous experiences and the problems presented with various *AI* efforts, the technology-centred perspective appears to be less promising, rather a waste of resources. In contrast, evidence-based examination reveals that the widespread use and secular success of computer technology is predominantly based on the praxeological *IA* perspective of *intelligence amplification* and the organizational development efforts connected to that approach. In this perspective, human skills, particularly reflective and conceptual learning capacities, are combined with the precision and velocity of the machine.

This must be put in the center of awareness in order to combine flexibility with efficiency. The socio-technical design then needs to be oriented at the peculiarities and needs of human acting and social practice. In particular, those conditions need to be regarded, under which human working capacity can unfold for increased productivity and creativity.

For accomplishment, working tasks sustaining competence and fostering learning, task-appropriate, transparent and controllable means of work with expected behaviour, as well as sufficient time resources for appropriating the tools and optimizing processes are needed (Brödner 2013). Four decades of extensive labour research provide a sound footing for that (although this knowledge base seems to presently fade away).

According to the dynamics of explicating practical skills as explicit codified knowledge and of appropriating this knowledge as augmented skill, the use of computer systems in organizations massively intervenes into their social practices, frequently with surprising results. Practical human skills and experiences supported by task-appropriate tools often prove to be superior to »smart« automatically operating systems replacing human actors; this is even true, if the automata perform better than human experts: The chess champion Kasparov e.g., who had been outperformed by IBM's »Deep Blue« computer, has on his part beaten again a comparably powerful computer by using a much simpler personal computer as a supportive tool (Kasparov 2010).

According to this type of human-computer interaction, continuously collected data from CPS might, for instance, be used as input in systems for interactive assistance with advanced usability to reconfigure or optimize production processes, to simulate and control such processes, or to use data analytics for preventive

maintenance. For effective interaction, it is important that users have opportunity to control the degree of detail with which they can look at the progress of machine or process states, at given settings, or at methods in use. This is needed in order to enable the users to generate a discrete picture of the incidents or the constitution of results, and to purposefully interfere. On the other hand, for handling the data safely, general agreements need to be accepted that regulate their use practice, particularly access conditions and operating methods.

To follow this praxeologically informed IA-perspective, means to accomplish higher flexibility, productivity, and innovation capacity by sociotechnical design of decent work, rather than betting on questionable AI-promises. It means to organize a productive, creative and autonomous cooperation of competent and knowledgeable experts supported by useful and usable computer artifacts such that their working capacity and competence can further grow. It lastly means to leave the road to subjection.

5 References

Anderson, C. (2008): The End of Theory, Wired 23.06.08

Bainbridge, L. (1983): Ironies of Automation, Automatica 19, 775-779

Bateson, G. (1980): Mind and Nature. A Necessary Unity. Toronto: Bantam Books

Bell, D. (1973): The Coming of the Post-Industrial Society. A Venture in Social Forecasting, New York: Basic Books

BMBF (2014): Innovationen für die Produktion, Dienstleistung und Arbeit von morgen, Bonn (Federal frame programme fostering innovation for manufacturing, services and work of tomorrow)

Boyd, D. (2011): Six Provocations for Big Data, www.softwarestudies.com/cultural_analytics/Six_Provocations_for_Big_Data.pdf [last access: 10.02.2015]

Breadshaw, J.M. (1997): An Introduction to Software Agents. In: Breadshaw, J.M. & Hutchinson, F. (eds.): Software Agents, Cambridge (MA): MIT Press, 3-46

Brödner, P. (2013): Reflective Design of Technology for Human Needs, 25th Anniversary Volume: A Faustian Exchange: What Is to Be Human in the Era of Ubiquitous Technology, AI & Society Journal of Human- Centred Systems 28, 27-37

Brödner, P. (2009): The Misery of Digital Organisations and the Semiotic Nature of IT, AI & Society Journal of Human-Centred Systems 23, 331-351

Brödner, P. (2007): From Taylorism to Competence-based Production, AI & Society Journal of Human-Centered Systems 21 (4), 497-514

Brödner, P. (1990): The Shape of Future Technology. The Anthropocentric Alternative, London: Springer

Brynjolfsson, E. & McAfee, A. (2014): The Second Machine Age. Work, Progress, and Prosperity in a Time ofBrilliant Technologies, New York London: Norton & Comp.

Carr, N. (2013): All Can Be Lost: The Risk of Putting Our Knowledge in the Hands of Machines, The AtlanticNo. 11

Dedrick, J.; Gurbaxani, V. & Kraemer, K.L. (2003): Information Technology and Economic Performance: A CriticalReview of the Empirical Evidence, ACM Computing Surveys 35, 1-28

Dennett, D.C. (1987): The Intentional Stance, Cambridge (MA): MIT Press

Drucker, P. F. (1994): The Age of Social Transformation, The Atlantic No. 11, 53-80

Eco, U. (1976): A Theory of Semiotics, Bloomington: Indiana University Press

Ehn, P., (1988): Work-Oriented Design of Computer Artifacts, Stockholm: Arbetslivscentrum

Floyd, C. (2002): Developing and Embedding Auto-operational Form, in: Dittrich, Y.; Floyd, C.; Klischewski, R. (Eds.): Social Thinking – Software Practice, Cambridge (MA): MIT Press, 5-28

Fodor, J. (1968): Psychological Explanation, New York: Random House

Foerster, H.v. (1991): Through the Eyes of the Other, in: Steyer, F. (Ed.): Research and Reflexivity, London: Sage Publications, 21-28

Giddens, A. (1984): The Constitution of Society, Outline of the Theory of Structuration, Cambridge: Polity Press

Gordon R.J. (2014): The Demise of US.Economic Growth: Restatement, Rebuttal, and Reflections, NBER Working Paper 19895

Hofstadter D. & Sander, E. (2013): Surfaces and Essences: Analogy as the Fuel and Fire of Thinking, New York: Basic Books

Hunt, V.D. (1989): Computer Integrated Manufacturing Handbook. London New York: Kluwer Academic Publishers

Jeschke, S. (2015): Auf dem Weg zu einer »neuen KI«: Verteilte intelligente Systeme, Informatik Spektrum 38 (1), S. 4-9 (Towards a »new AI«: Distributed intelligent systems)

Jorgenson, D. W.; Ho, M. S. & Stiroh, K. J. (2008): A Retrospective Look at the U.S. Productivity Growth Resurgence, Journal of Economic Perspectives 22 (1), 3-24

Kasparov, G. (2010): The Chess Master and the Computer, The New York Review of Books 11.02.2010

Kriesel, D. (2007): A Brief Introduction to Neural Networks, http://www.dkriesel.com/science/neural_networks [last access: 12.02.2017]

Maes, P. (1994): Agents that Reduce Work and Information Overload, CACM 37 (7), 31-41

Minsky, M. (1988): The Society of Mind, New York: Simon & Schuster

Nonaka, I., (1996): A Dynamic Theory of Organizational Knowledge Creation, Organization Science 5 (1), 14-37

Norman, D. A. (1994): How Might People Interact with Agents, CACM 37 (7), 68-71

Norman, D. A. (1993): Things that Make Us Smart, Reading (MA): Addison-Wesley

Peirce, C. S. (1903): A Syllabus of Certain Topics of Logic, Collected Papers, 1.180-202, 2.219-225 and other paragraphs

Perrow, C. (1984): Normal Accidents. Living with High-Risk Technologies, New York: Basic Books

Polanyi, M. (1966): The Tacit Dimension, Garden City (NY): Doubleday

Prasse, M. & Rittgen, P. 1998: Bemerkungen zu Peter Wegners Ausführungen über Interaktion und Berechenbarkeit, Informatik-Spektrum 21, 141-146 (Remarks on Peter Wegner's statements about interaction and computability)

Putnam, H. (1960): Minds and Machines. In: Hook, S. (ed.): Dimensions of Mind, New York: Collier Books

Putnam, H. (1991): Representation and Reality, Cambridge (MA): MIT Press

Reckwitz, A. (2002): Toward a Theory of Social Practices: A Development in Culturalist Theorizing, European Journal of Social Theory 5 (2), 243-263

Ryle, G. (1949): The Concept of Mind, London: Hutchinson

Schmidhuber, J. (2015): Deep Learning in Neural Networks. An Overview, Neural Networks 61, 85–117

Searle, J. R. (2010): Making the Social World. The Structure of Human Civilization, Oxford: Oxford University Press

Shannon, C. (1948): A Mathematical Theory of Communication, The Bell Systems Technical Journal 27, 379-423 and 623-656

Sharif, M. et al. (2016): Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition, ACM Conference on Computer and Communication Security, Vienna

Shoham, Y. (1993): Agent-Oriented Programming, Artificial Intelligence 60, 51-92

Solow, R. (1987): We We'd Better Watch Out, Th New York Times Book Review, July 12

Spur, G. (1984): Über intelligente Maschinen und die Zukunft der Fabrik, Forschung – Mitteilungen der DFG I-VIII (About intelligent machines and the future of the factory)

Varela, F. J.; Thompson, E. & Rosch, E. (1991): The Embodied Mind. Cognitive Science and Human Experience. Cambridge (MA): MIT Press

WEF (2012): The Future of Manufacturing. Opportunities to drive economic growth, a World Economic Forum Report in collaboration with Deloitte Touche Tohmatsu Limited

Wegner, P. (1997): Why Interaction is More Powerful than Algorithms, CACM 40 (5), 80-91

Wiener, N. (1950): The Human Use of Human Beings. Cybernetics and Society, Boston (MA): Houghton Mifflin Harcourt

Winograd, T. (1996): Bringing Design to Software, Reading (MA): Addison-Wesley

Wooldridge, M. (2002): An Introduction to Multi-Agent Systems, New York: Wiley

↑Inhalt↑

Industrie 4.0 und Big Data – wirklich ein neuer Technologieschub?

1 Einführung: Der Traum von der automatischen Fabrik

Derzeit erleben wir unter den Parolen »Industrie 4.0 und Big Data« eine neue Welle technologischen Überschwangs, in der fortgeschrittene Computertechnik geradezu als Heilsbringer verehrt wird. Kaum ein Übel der Lebenswelt erscheint zu groß, als dass nicht dessen Überwindung mittels »Digitalisierung« in Aussicht gestellt würde (Forschungsunion & acatec 2013, S. 5). So bietet das Erscheinen der deutschen Übersetzung des Buches von Brynjolfsson & McAfee (2014) »The Second Machine Age. Wie die nächste digitale Revolution unser aller Leben verändern wird« und des neuen Rahmenprogramms »Innovationen für die Produktion, Dienstleistung und Arbeit von morgen« des BMBF (2014) genügend Anlass, anhand dort ausgeführter Überlegungen zur Sicherung künftiger Wertschöpfung die These von einer neuen Welle technologischen Überschwangs exemplarisch zu untermauern, Hintergründe aufzuzeigen und Aussichten zu bewerten.

In ihrem Buch begründen Brynjolfsson und McAfee (2014) das Heraufziehen eines neuen Maschinenzeitalters mit der außerordentlichen, auf exponentiellem Wachstum beruhenden Steigerung von Leistungen der »Digitaltechnik« (Computer und Netzwerke; »Mooresches Gesetz«), mit den dadurch ermöglichten raschen Fortschritten bei der Realisierung »künstlicher Intelligenz« und der Verarbeitung sehr großer Datenbestände im Zuge weitreichender Digitalisierung von Wertschöpfung, zudem mit unbegrenzter (Re-)Kombination von Ideen zur Innovation. Sie exemplifizieren das etwa an den Fällen des selbstfahrenden Autos von Google oder der Wissensverarbeitung durch IBM Watson. Allerdings zeigen sie auch Gefahren der Monopolisierung globaler digitaler Wertschöpfung durch Netzwerkeffekte und geringe Grenzkosten (sog. »Alles-oder-nichts-Märkte«), verweisen auf die Polarisierung von Qualifikation und Einkommen und diskutieren Möglichkeiten steuernder Eingriffe in die künftige Entwicklung.

Das BMBF-Programm »Innovationen für die Produktion, Dienstleistung und Arbeit von morgen« setzt – als zentraler Bestandteil der »High-Tech«-Strategie der Bundesregierung zum Erhalt von Wettbewerbsfähigkeit und Wohlstand – ebenfalls auf Digitaltechnik, insbesondere auf die Entwicklung weltweit vernetzter, sog. »Cyber-Physical Systems« (CPS). Als Kernkomponenten von Produktion und Logistik werden dabei »intelligente« Maschinen, Werkstücke, Lagersysteme und Betriebsmittel ins Auge gefasst, die mittels fortgeschrittener Computertechnik

befähigt werden sollen, in »dezentraler Selbstorganisation« selbsttätig Daten auszutauschen, gegenseitig Aktionen auszulösen und so eine weitgehend flexibel automatisierte »Smart Factory« zu realisieren. Im Einklang mit entwicklungsstrategischen Überlegungen des World Economic Forum (WEF) wird dabei verheißen, neue Formen von Wertschöpfungsprozessen – unter Berücksichtigung individueller Kundenwünsche – anpassungsfähig und dynamisch gestalten, auch Einzelstücke rentabel herstellen, flexibel auf Störungen und Ausfälle reagieren, durchgängig Transparenz gewährleisten und optimale Entscheidungen ermöglichen zu können (BMBF 2014, Forschungsunion & acatec 2013, WEF 2012).

In vielerlei Hinsicht ruft das Erinnerungen an frühere Versuche wach, mittels Computer- und Sensortechnik den wiederkehrenden Traum von der automatischen Fabrik, von ewiger selbsttätiger Wertschöpfung zu verwirklichen und so vom Eigensinn lebendiger Arbeit unabhängig zu werden. Schon in den frühen 1950er Jahren hatte Norbert Wiener (1950/66, 167f) sehr genaue Vorstellungen davon, wie mittels »Sinnesorganen« und »Effektoren« sowie der »Rechenmaschine« als »zentralem logischen Gehirn« für »komplizierte Unterscheidungsprozesse« eine »automatische Fabrik« zu realisieren sei. Und in den 1980er Jahren wurden unter dem Leitstern der »menschenleeren Fabrik« erneut gigantische F&E-Anstrengungen unternommen, um in Gestalt »wissensbasierter« (d.h. mit symbolischer »künstlicher Intelligenz« ausgestatteter), »computer-integrierter Produktion« (CIM) eine flexibel automatisierte auftragsgebundene Fertigung zu verwirklichen (Cyranek & Ulich 1993, Hunt 1989). Diese sich mit einer Länge von rd. 30 Jahren auftürmenden Flutwellen technologischen Überschwangs brachen sich jedesmal an den Klippen widerspenstiger Materie und verkannter Realisierungsprobleme mit der Folge langer Phasen tiefer Ernüchterung bei deren Bewältigung. Ironischerweise wurde dabei regelmäßig der hohe Wert impliziten Wissens, der Intuition, Kreativität und Handlungskompetenz menschlichen Könnens wiederentdeckt.

Um vor diesem Hintergrund Ähnlichkeiten und Unterschiede zu den neuen, doch sehr weit reichenden Verheißungen von »Industrie 4.0 und Big Data« realistisch einschätzen zu können, will der Beitrag zunächst deren wissenschaftlich-technische Grundlagen näher kennzeichnen und das eigentlich Neue im Vergleich zu früheren Ansätzen genauer bestimmen. Darauf fussend wird dann eine kritische Bewertung von Entwicklungs- und Gefahrenpotenzialen vorgenommen, um abschließend Gestaltungsperspektiven einer zukunftsfähigen Entwicklung zu skizzieren.

2 Worum geht es? – Kennzeichnung von Industrie 4.0 und Big Data

2.1 Schwierigkeiten der Bestimmung des gegenwärtigen Wandels

Mit den plakativen Benennungen »Industrie 4.0« (BMFT 2014) bzw. »zweites Maschinenzeitalter« (Brynjolfsson & McAfee 2014) soll jeweils ein qualitativer Sprung in der industriellen Entwicklung im Vergleich zu vorangegangenen Veränderungen markiert werden. So sei aus der Perspektive Industrie 4.0 die »vierte industrielle Revolution« durch den Einsatz von stark vernetzten »cyber-physischen Systemen« gekennzeichnet im Unterschied zur computergestützten Automatisierung der Produktion der dritten Stufe. Diese wiederum unterscheide sich von der durch Massenfertigung und elektrische Einzelantriebe an Fertigungsanlagen gekennzeichneten zweiten wie auch von der durch Mechanisierung und Dampfantrieb charakterisierten ersten industriellen Revolution (Forschungsunion & acatec 2013). Ähnlich wird auch das durch den forcierten Einsatz von avancierter Digitaltechnik (Computer und Netzwerke) gekennzeichnete zweite Maschinenzeitalter (etwa den Stufen 3 und 4 entsprechend) von der Mechanisierung und Elektrifizierung der Produktion als Kennzeichen des ersten (etwa den Stufen 1 und 2 entsprechend) abgegrenzt (Brynjolfsson & McAfee 2014).

Derartige technikzentrierte Unterscheidungen von Phasen gesellschaftlicher Entwicklung erweisen sich freilich eher als irreführend denn erhellend, ganz abgesehen davon, dass sie mindestens ebenso bedeutsame soziale, organisationale und institutionelle Aspekte ausblenden wie etwa die betriebliche Arbeitsteilung, die Standardisierung oder die Wissensteilung. Damit geraten viel tiefer gehende Prozesse gesellschaftlichen Wandels wie der Übergang von der Industrie- zur Wissensgesellschaft (Bell 1975, Drucker 1994) mit seinen erhöhten Anforderungen an das lebendige Arbeitsvermögen, insbesondere die Fähigkeit zu produktiver Kooperation bei der Genese, Organisation und Anwendung von Wissen, ganz aus dem Blick.

Dabei ist für das Verständnis wissensintensiver Arbeit das Verhältnis von praktischem Können und kodifiziertem Wissen wesentlich, genauer: deren dynamische Beziehung, die Art und Weise, wie sie einander wechselseitig hervorbringen. Vorgängig ist stets die vorreflexive Handlungskompetenz, das individuell gebundene Können oder Arbeitsvermögen, das sich in Tätigkeiten gelingender sozialer Praxis äußert. Bei einer – wie auch immer – gestörten Praxis lässt sich durch Selbst- oder Fremdbeobachtung explizites, theoretisches Wissen *über* bestimmte Aspekte praktischen Tätigseins gewinnen. Dieses kodifizierte Wissen bleibt aber ohne Wirkung, solange es nicht zur Verwendung in praktischer Tätigkeit angeeignet und zweckmäßig genutzt wird. Gleiches gilt auch für solches Wissen verkörpernde technische Artefakte (Brödner 2010; vgl. allgemein zu dieser praxistheoretischen Perspektive Reckwitz 2003, bezüglich Computersystemen Brödner 2008).

Dieser Dialektik der – stets begrenzten – Explikation von Können und Erfahrung als Wissen und der Aneignung von Wissen (bzw. technischen Artefakten) als erweitertem Können zufolge erfordert produktives Arbeiten und Problemlösen in Prozessen wissensintensiver Wertschöpfung meist, diverse Wissensgebiete zusammenzuführen und unterschiedlich ausgeprägte Arbeitsvermögen in Gestalt autonomer, selbstorganisierter Kooperation kompetenter Experten zu integrieren. Je differenzierter, komplexer und dynamischer das kodifizierte Produktionswissen und dessen technische Vergegenständlichung, desto anspruchsvolleres Arbeitsvermögen ist gefordert, sich dieser gesellschaftlichen Produktivkräfte zu produktiver Verwendung zu bemächtigen und sie weiter zu entwickeln. Das funktioniert nur freiwillig, entzieht sich jeder Anweisung und unterliegt zudem großer Ungewissheit hinsichtlich Ergebnis und Verlauf.

Um das Primat der Kapitalverwertung auch im unsicheren Umgang mit Wissen in weitgehend autonomen, sich selbst organisierenden Arbeitsgruppen zu gewährleisten, sind neue Formen der Kontrolle gefragt: *indirekte Steuerung* durch Markt- oder Kontextanforderungen anstelle hierarchischer Weisung und Kontrolle. Darin sehen sich die Menschen als Träger des Arbeitsvermögens permanentem Erfolgsdruck ausgesetzt und treiben sich selbst unablässig zu Höchstleistungen an, allerdings auf Kosten ihrer Gesundheit und sozialen Beziehungen. Das immer wichtiger werdende Arbeitsvermögen kann sich mithin unter diesen Bedingungen seiner Verausgabung in den Arbeitsprozessen selbst nicht hinreichend entfalten (Peters & Sauer 2006).

Schließlich sind in diesem Kontext noch sehr grundlegende Unterschiede zwischen Computern als zeichenverarbeitenden Maschinen und klassischen Maschinen der Energie und Stoffumwandlung zu beachten und zwar im Hinblick auf deren Wirkbereiche, Funktionsweisen und Zwecke: Der Wirkbereich von Arbeits- und Kraftmaschinen liegt in der Natur und greift in natürliche Prozesse der Energie- und Stoffumwandlung ein, während der Wirkbereich von Computern ganz im Bereich von Zeichenprozessen sozialer Praxis liegt und auf die algorithmische Verarbeitung damit verbundener Signale oder Daten zielt. Dementsprechend beruht die Funktionsweise von Maschinen der Energie- und Stoffumwandlung auf Natur-Effekten als Ergebnis von Naturerkenntnis und ihr Zweck ist die Nutzung von Naturkräften als Quelle von Produktivität. Die Funktionsweise von Computern beruht dagegen auf expliziten, durch Analyse und Formalisierung von Zeichenprozessen gewonnenen Vorschriften (Algorithmen). Sie dient entweder der digitalen Steuerung technischer Prozesse (»eingebettete Systeme«) oder der Organisation und Koordination kollektiven Handelns (»Informationssysteme«). Bei letzterem kann höhere Produktivität nur aus der Art und Weise erwachsen, wie Zeichenprozesse sozialer Praxis durch passend gestaltete Computerfunktionen und deren kollektive Aneignung neu strukturiert und organisiert werden.

2.2 Eingebettete und Multiagenten-Systeme als wissenschaftlich-technisches Fundament

Die Realisierung von Industrie 4.0 stützt sich auf wissenschaftlich-technische Grundlagen, deren Entwicklung seit den 1990er Jahren massiv gefördert und vorangetrieben wird:

- Eingebettete Systeme sind Computer-Komponenten zur digitalen Steuerung physischer Prozesse, die mit Schnittstellen zum Menschen und anderen Systemen ausgestattet sind. Sie können über Datenaustausch durch das Internet hochgradig horizontal und vertikal vernetzt werden („Internet der Dinge & Dienste“ bzw. „Cyber-Physical Systems“; vgl. Abb. 1a und b).

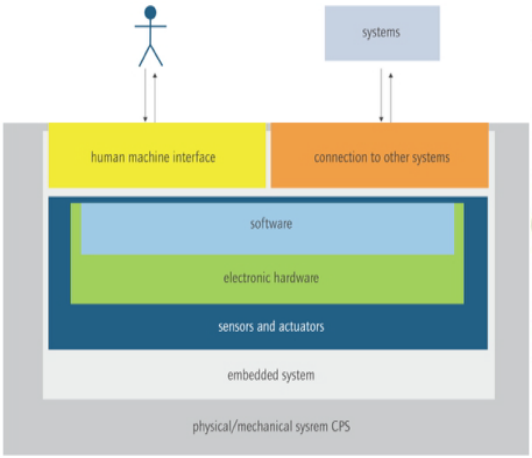


Abb. 1a: *Eingebettetes System*

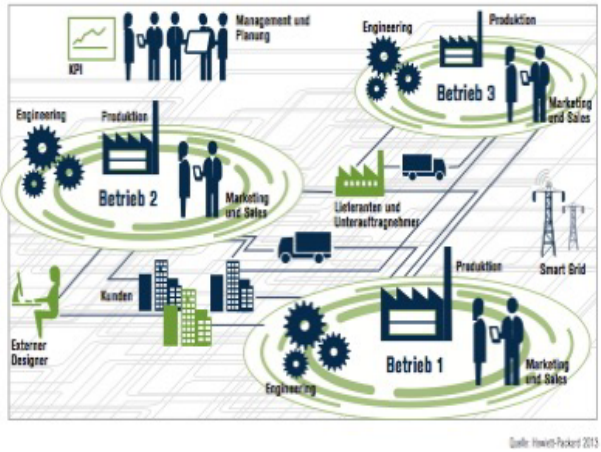


Abb. 1b: *Cyber-Physical Systems in der Produktion*

- Multiagentensysteme (MAS‚auch: »Distributed Artificial Intelligence«) bestehen aus begrenzt autonomen Software-Agenten, die mittels Datenaustausch zielorientiert miteinander interagieren; durch derart koordinierte Aktionen sollen sie anspruchsvolle Aufgaben selbsttätig bewältigen können.

Zum besseren Verständnis des weiteren ist ein genauerer Blick auf die Funktionsweise, das Verhalten und die Beschreibung von MAS erforderlich. Das außerordentliche Interesse an MAS beruht auf der Vorstellung, aus dem Zusammenwirken vieler Einheiten mit relativ einfachem Verhalten könne »künstliche Intelligenz« hervorgehen (Minsky 1988). Sie gründet sich letztlich auf die explizit artikulierte Überzeugung, dass »Interaktion mächtiger als Algorithmen« sei (Wegner 1997). Letzteres wurde freilich postwendend als Trugschluss widerlegt (auch MAS unterliegen der Beschränkung auf berechenbare Funktionen; Prasse & Rittgen 1998). Tatsächlich weist aber das Verhalten von MAS emergente Eigenschaften auf, die an den einzelnen darin zusammenwirkenden Software- Agenten für sich nicht zu beobachten sind.

Agenten sind softwaretechnische Objekte, die Daten aus der Umgebung und von anderen Agenten aufnehmen, proaktiv nach eigenen Algorithmen, zumeist

mittels maschineller Lernverfahren, verarbeiten und resultierende Daten wieder nach außen abgeben können. Ihr Verhalten ist gekennzeichnet durch selbsttätige Zielverfolgung und die Fähigkeit zur Anpassung an veränderliche Bedingungen durch »maschinelles Lernen« (Breadshaw 1997, Maes 1994). Zwecks Bewältigung komplexer Aufgaben können solche begrenzt autonomen Software-Agenten miteinander interagieren, um derartige Aufgaben durch koordinierte Aktionen gemeinsam zu erledigen (Wooldridge 2002). Zudem fallen dabei laufend große, noch gesondert nutzbare Datenmengen an.

Auch wenn einzelne im Systemverbund interagierende Agenten relativ einfache Algorithmen ausführen und gut durchschaubares Verhalten aufweisen, ist dem MAS insgesamt zwar ein vollständig determiniertes, aber hoch komplexes und analytisch von außen nicht mehr bestimmbares Verhalten eigen. Formal lassen sich MAS als sog. »nichttriviale Maschinen« (v. Foerster 1993) beschreiben, deren Ausgabedaten nicht nur von den Eingabedaten, sondern gemäß einer Zustandsfunktion auch von veränderlichen inneren Zuständen abhängen, die auf vielfältige Weise die Interaktion der Agenten und deren Lernverhalten zum Ausdruck bringen (vgl. Abb. 2). Folglich ist das Verhalten von MAS in hohem Maße von der jeweiligen Vorgeschichte abhängig, analytisch nicht mehr bestimmbar und mithin auch nicht vorhersehbar.

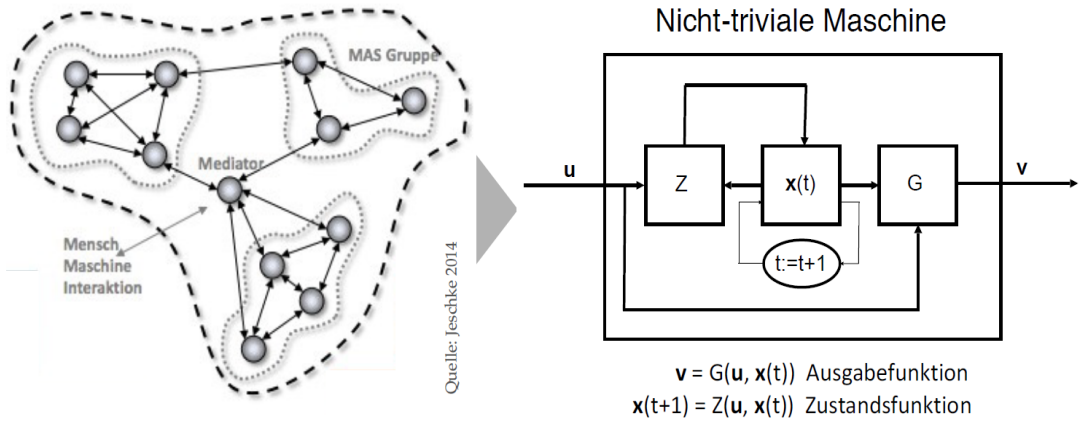


Abb. 2: *MAS als nicht-triviale Maschine*

Neben dieser physikalisch-algorithmischen Beschreibung des Verhaltens von MAS ist noch eine weitere, eher an der zielgerichteten Interaktion der Agenten orientierte Beschreibung gemäß der sog. »intentionalen Einstellung« (»*intentional stance*«, Dennett 1987) üblich. Dieser Einstellung zufolge werden Agenten intentionale Zustände wie Überzeugungen, Wünsche oder Absichten zugeschrieben, ausdrücklich um deren Verhalten einfacher und übersichtlicher beschreiben zu können; sie wird unter MASForschern allgemein geteilt und geht auf einen von McCarthy verfassten Forschungsbericht aus dem Jahre 1978 zurück:

> *»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. Theories of belief, knowledge and wanting can be constructed for machines in a simpler setting than for humans, and later applied to humans. Ascription of mental qualities is most straightforward for machines of known structure such as thermostats and computer operating systems, but is most useful when applied to entities whose structure is incompletely known.« (*McCarthy 1978, zitiert nach Shoham 1993, S. 53*).*

Zur formalen Beschreibung intentionaler Zustände von Agenten bedient man sich der propositionalen Modallogik, in der die übliche zweiwertige Aussagenlogik (eine Aussage ist entweder wahr oder falsch) um die beiden Modalitäten der Notwendigkeit und der Möglichkeit erweitert ist: Eine Aussage ist zudem notwendiger- oder aber nur möglicherweise wahr oder falsch. Auf dieser erweiterten logischen Grundlage lassen sich dann Überzeugungen (*beliefs*), Wünsche (*desires*) oder Absichten (*intentions*) von Software- Agenten als zugeschriebene intentionale Zustände formalisieren; beispielsweise kann so die Tatsache, dass der Agent a zum Zeitpunkt t der Überzeugung ist, dass die Aussage ϕ gilt, als der Ausdruck $\mathbf{Bel}(a, \phi)(t)$ einer formalen Beschreibungssprache modelliert werden (Wooldridge 2002).

Diese durch Zuschreibung von Intentionalität gewonnene Abstraktion vom wirklichen, auf der physikalisch-algorithmischen Ebene sich vollziehenden Verhalten der MAS ist ein Versuch, dem ärgerlichen Umstand zu entkommen, dass dieses Verhalten in seinem physischen Verlauf, obgleich determiniert, i.a. nicht mehr zu durchschauen ist. Weil das Verhalten auf dieser Ebene nicht erklärbar ist, wird in der Abstraktion so getan, als ob ihm Intentionalität zugrunde läge. Diese letztlich absurde *Quasi*-Erklärung des Verhaltens von MAS gaukelt dessen Verstehbarkeit vor, der das wirkliche Verhalten tatsächlich aber entzogen bleibt; statt das Problem zu erhellen oder zu lösen, wird es verschleiert. Beides, die mangelnde Durchschaubarkeit des Verhaltens von MAS wie die Zuschreibung von Intentionalität als dessen *Quasi*-Erklärung, zeitigt aber, wie noch zu zeigen ist, fatale Konsequenzen für die Mensch-Maschine-Interaktion und die Sicherheit von MAS (Norman 1994).

2.3 »Maschinelles Lernen« als neuer Königsweg?

Ein inhärenter Nachteil algorithmisch gesteuerten Verhaltens ist dessen Starrheit, die sich nur in Interaktion mit lebendiger Arbeit oder durch eigene Adaptivität mittels Verfahren »maschinellen Lernens« überwinden lässt. Diese umfassen ein breites Spektrum von Methoden und ihrerseits algorithmischen Verfahren zur ge-

zielten Veränderung der Struktur und Parameter bzw. der Programme von Computern dergestalt, dass sich deren Verhalten gemessen an einer zuvor festgelegten Nutzenfunktion verbessert. Benötigt werden sie dort, wo Computer anspruchsvolle Aufgaben selbsttätig und zielorientiert zu bewältigen oder ihr Verhalten an wechselnde Umgebungsbedingungen anzupassen haben.

Durch aufsehenerregende Leistungen haben in jüngster Zeit sog. künstliche neuronale Netze (KNN) und Verfahren des »Deep Learning« (Wick 2017) große Aufmerksamkeit gefunden. Von Beginn an mit der Computertechnik eng verbunden, erfahren sie derzeit, nach dem von Vertretern der »symbolischen KI« ausgesprochenen Verdikt und einer gewissen Wiederbelebung in den 1990er Jahren (Brödner 1997, S. 206ff), eine abermalige Renaissance. Diese verdankt sich vor allem exponentiell gesteigerter Rechenleistung, die komplexere Netzwerk-Strukturen und erheblich beschleunigte Lernverfahren ermöglicht.

Für ihre Aufgaben müssen KNN passend strukturiert und die Gewichtungen der in ihre als »Neuronen« fungierenden Prozessoren eingehenden Signale einem bestimmten Lernalgorithmus unterworfen werden. Dabei berechnet jeder Prozessor (Netzknoten) *j* die in Bild 3a dargestellten Funktionen und liefert ein Aktivierungssignal, falls die Summe der gewichteten Eingabesignale einen Schwellwert θj überschreitet. Das Netz als Ganzes wird über diese Festlegungen und Lerninstruktionen hinaus nicht programmiert, sondern an sehr vielen Beispielen trainiert (zur Einführung in KNN vgl. Kriesel o.J.).

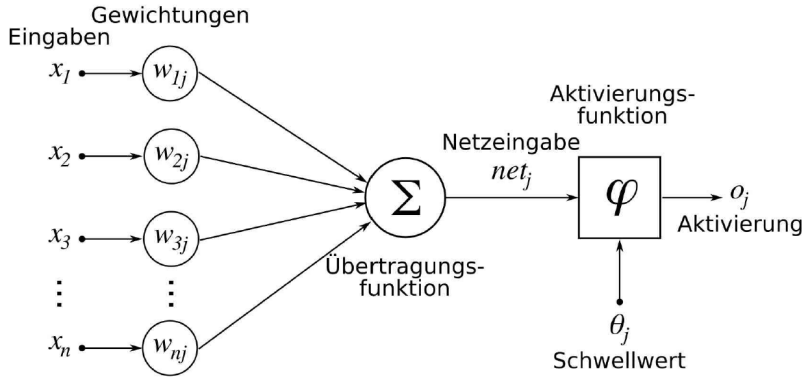


Bild 3a: *Berechnungsfunktionen eines Netzknotens*
(Quelle: Wikipedia CC BY-SA 3.0)

So werden etwa bei Problemen der Musteridentifikation oder der (Bild-)Klassifikation – ein Aufgabentyp, bei dem KNN besonders leistungsfähig sind – einer langen Reihe von eingegebenen Mustern (Bildern) jeweils die zugehörigen Klassen als Ausgänge zugeordnet; aus diesen Zuordnungen vermag dann das Netz automatisch in zahlreichen kleinen Anpassungsschritten angemessene Verbindungsgewichte w_{ij} zu bestimmen (die implizit den Lernerfolg widerspiegeln). Dabei werden je Zwischenschicht von Prozessoren bzw. »Neuronen« spezifische Filter wirksam, die sukzessive kennzeichnende Merkmale aus den Bildern zu ex-

trahieren in der Lage sind. Am Ende wird in der Ausgabeschicht jeweils genau ein Prozessor aktiviert, der jeweils der errechneten Kategorie entspricht (im Beispiel etwa ein Auto oder eine Gesicht; vgl. Bild 3b).

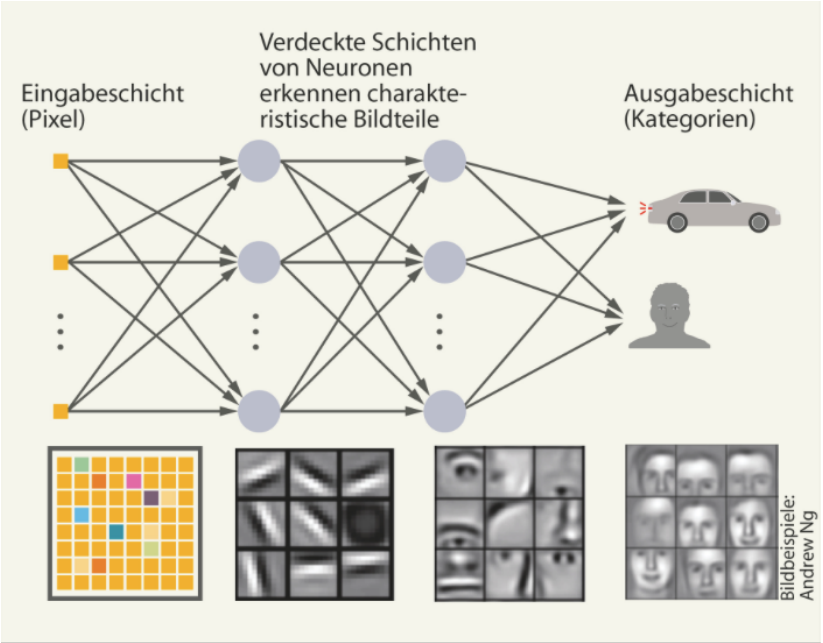


Bild 3b: *Kategorisierung von Bildern (Quelle: c't 6/2016)*

Dementsprechend verdanken sich Erfolge mit dem Einsatz von KNN nicht tieferer Einsicht in Zusammenhänge von Ursache und Wirkung, sondern weitgehend einem theorielosen Probieren durch Versuch und Irrtum mit unterschiedlichen Netzwerk- Strukturen und Lernverfahren zur Anpassung der Gewichte. So sind derzeit wegen ihrer besonderen Leistungsfähigkeit bei Bild- und Spracherkennungsaufgaben vor allem sog. »faltende« KNN (Convolutional Neural Networks) ins Zentrum des Interesses gerückt, die sich strukturell an biologische Vorbilder anlehnen. Während konzeptionell kaum Fortschritte zu verzeichnen sind, beruhen eingetretene Erfolge neben der gesteigerten Rechenleistung ironischerweise vor allem auf der Intuition, der Erfahrung und dem Geschick der Entwickler bei der Strukturierung der KNN und der Bewältigung der zahlreichen praktischen Berechnungsprobleme – vom »schwindenden Gradienten« bis zur variablen Schrittweitensteuerung bei den Lernverfahren (Schmidhuber 2015, Wick 2017).

Beim Gebrauch dieser Systeme im praktischen Einsatz haben die Nutzer zudem mit der Ungewissheit zu kämpfen, ob die errechneten Ergebnisse auf Dauer angemessen und korrekt sind. Auch wenn sie in der großen Mehrzahl der Fälle gute Ergebnisse liefern, können sie plötzlich versagen, ohne dass das einfach erkennbar wäre. Zudem können schon kleine Störungen in den Eingabebildern zu beträchtlichen Fehlleistungen führen (Sharif et al. 2016). Mangels Durchschaubarkeit des nicht-linearen Verhaltens von KNN ist deren Zuverlässigkeit generell

schwer zu beurteilen; im Grunde müssen sich die Nutzer daher blindlings auf die angemessene Funktionsweise verlassen.

2.4 Big Data: Fast unerschöpfliche Datenbestände im Zugriff

Big Data – Unter dieser Bezeichnung wird üblicherweise die schnelle Analyse sehr großer und vielfältiger, mehr oder weniger strukturierter Datenbestände für Aktionen, Planungen und Prognosen verstanden. Vielfältige Datenbestände können aus unterschiedlichen Quellen stammen, verschiedene Grade der Strukturierung aufweisen (Bild- oder Textdokumente, Datenbanken) und gleichwohl miteinander kombiniert werden. Weitere Kennzeichen sind hohe Datenvolumina (im Bereich von Tera- oder gar Petabyte) und große Verarbeitungsgeschwindigkeiten (mehr oder weniger in Echtzeit).

Aufgrund der enormen Leistungssteigerung von Computerprozessoren und -speichern ist es möglich geworden, sehr große Datenmengen im Arbeitsspeicher bereitzuhalten und damit in sehr schnellem Zugriff algorithmisch zu verarbeiten (sog. »In-memory-Technik«). Die Ausweitung von Bandbreiten erlaubt auch die sehr schnelle Übertragung großer Datenmengen. Zudem können, soweit es die Aufgabenstellung zulässt, Datenbestände aufgeteilt und in einer Vielzahl von Prozessoren parallel verarbeitet werden. So können die exponentiell gesteigerten Leistungspotenziale der Digitaltechnik unmittelbar ausgespielt und zur Bewältigung komplexer Aufgaben genutzt werden.

3 Was ist neu? – Vergleich mit früheren Ansätzen

Benennungen wie »vierte industrielle Revolution« oder »zweites Maschinenzeitalter« suggerieren einen grundlegenden Entwicklungssprung von erheblicher Tragweite. Genau dies lassen aber die vorgestellten neuen Ansätze vor dem Hintergrund der tatsächlichen Entwicklung der letzten drei Dekaden als fraglich erscheinen.

Erklärtes Ziel der neuen Ansätze ist es, Wertschöpfungsprozesse anpassungsfähig zu gestalten, auch Einzelleistungen rentabel zu produzieren und auf Störungen flexibel zu reagieren. Das sind aber genau die gleichen Anforderungen, die schon in den 1980er Jahren durch computer-integrierte und wissensbasierte Produktion (CIM) erreicht werden sollten. Heute wie damals beherrscht die technikzentrierte Sicht auf Produktion das Feld, gibt es eine Welle technikzentrierten Überschwangs und untaugliche Versuche, Probleme der Organisation von Produktionsprozessen technisch zu bewältigen.

Auch bei CIM ging es bereits darum, möglichst viele Komponenten rechnerunterstützter Fertigung zu vernetzen und Daten zwischen ihnen auszutauschen. Derart vernetzte Systeme wurden in der Folge auch auf vielfache Weise realisiert, allerdings wurden deren Funktionen meist, anders als ursprünglich gedacht, auf Veranlassung durch und in Interaktion mit menschlichen Experten und deren

Arbeitsvermögen genutzt. Dagegen sind Versuche der Realisierung einer weitgehenden flexiblen Automatisierung mittels Expertensystemen und anderen wissensbasierten Systemen kläglich gescheitert. Stattdessen haben sich in breiter Front zellular organisierte Arbeitsstrukturen entwickelt mit hoch qualifizierten, sich weitgehend selbst steuernden Arbeitsgruppen unterschiedlicher Experten, jeweils unterstützt durch gebrauchstauglich gestaltete Computerfunktionen. Derartige »High-road«- oder »High-performance«-Organisationen ermöglichen die Entfaltung von Arbeitsvermögen und haben sich als hoch produktiv und innovativ erwiesen (Brödner 2006) – ganz im Einklang mit der ressourcenbasierten Sicht auf Unternehmen (Barney 1991, Grant 1996, Penrose 1995).

Statt wissensbasierter Expertensysteme sollen nun also im neuen Anlauf zur »Smart Factory« bei gleicher Zielsetzung Multiagentensysteme in Gestalt »lernfähiger«, hochgradig vernetzter »cyber-physischer Systeme« ins Rennen geschickt werden, um die äußerst verwickelten Koordinationsaufgaben numerisch und funktional flexibler, auftragsgebundener Wertschöpfung möglichst weitgehend selbsttätig zu bewältigen. Wie weit dieser neue Ansatz trägt, wird die Zukunft erweisen; Erfahrungen mit früheren Versuchen und Konzepten »künstlicher Intelligenz« lassen das aber als höchst zweifelhaft erscheinen (Brödner 1997).

Das abermalige Vorherrschen einer technikzentrierten Sicht vermittelt einen starken Eindruck von Déjà-vu; er deutet darauf hin, dass aus Problemen früheren technikzentrierten Einsatzes und Gebrauchs komplexer Computersysteme in Wertschöpfungsprozessen wenig gelernt wurde. Dagegen steht freilich das Fazit aus 25 Jahren Forschung über Produktivität und Computereinsatz in Organisationen, derzufolge Produktivität – aus praxistheoretischer Sicht erwartungsgemäß – nur aus der Kombination mit organisationaler Restrukturierung, Aneignung und Lernen erwächst:

> *»To leverage information technology investments successfully, firms must typi cally make large complementary investments and innovations in areas such as business organization, workplace practices, human capital, and intangible capital.«* (Jorgenson et al. 2008, S. 10; ebenso bereits Dedrick et al. 2003).

Schließlich lässt die genauere Analyse der technischen Konzepte von Industrie 4.0 selbst an deren Neuigkeitswert zweifeln. Was unter der Bezeichnung »cyber-physische Systeme« propagiert wird, ist zunächst nichts anderes als die Fortführung dessen, was seit den 1970er Jahren als digitale Prozesssteuerung intensiv betrieben wird. Neu ist hier lediglich die mit Erweiterung des Adressraums gegebene Möglichkeit, über das »Internet der Dinge« Daten zwischen sehr vielen digital gesteuerten Prozessen selbsttätig austauschen zu können, was Möglichkeiten der Reorganisation von Wertschöpfung erweitert (aber auch Risiken erhöht). Zudem lässt die seit über zwei Dekaden betriebene MAS-Forschung kaum praxisrelevante Fortschritte erkennen. Höchst wirksam zeigt sich dagegen die exponentielle Leistungssteigerung der Digitaltechnik hinsichtlich Rechenleistung, Speicherkapazität und Bandbreite der Datenübertragung. So entpuppt sich

die »vierte industrielle Revolution« vor allem als eine Revolution der Worte (»CPS« und »MAS« statt »digitaler Steuerung«), bei freilich enorm gesteigerter Leistung der Digitaltechnik, die früher außer Reichweite liegende Anwendungen neuerdings möglich macht.

4 Wozu dient es? – Kritische Bewertung der technischen Basis

4.1 Brüchige Fundamente der Entwicklung von MAS

So wird die neue Technikeuphorie gespeist durch enorm gesteigerte Computerleistungen und einzelne Beispiele fortgeschrittener Implementierung menschenähnlicher Fähigkeiten wie etwa das selbstfahrende Auto von Google oder die Wissensmaschine Watson von IBM, auf die sich die Botschafter dieses heilsbringenden Technologieschubs zum Nachweis seiner Realitätsnähe gerne berufen, so auch Brynjolfsson und McAfee (2014); sie sind in der Tat beeindruckend und liefern verblüffende Resultate.

Deren Überzeugungskraft schwindet aber rasch wieder, wenn man sich die extrem hohen Entwicklungsaufwendungen (in der Größenordnung von 103 Personenjahren) vor Augen führt, auf denen diese Leistungen beruhen und die jeweils ganz speziell auf die zu bewältigenden Aufgaben zugeschnitten sind: u.a. eine in jedem Detail sehr genaue dreidimensionale Kartierung der Fahrwege (einschl. ihrer – veränderlichen – Begrenzungen) im Falle des selbstfahrenden Autos oder die anwendungsspezifische Implementierung von Heuristiken und sehr großer Bestände enzyklopädischen Wissens im Falle von Watson. Zwar können grundlegende Verfahren etwa der Bildverarbeitung, der statistischen Analyse oder logischer Schlussweisen aufgabenübergreifend wiederverwendet werden, gleichwohl bleiben zum Erschließen neuer Einsatzfälle jeweils aufwendige aufgaben- und kontextspezifische Entwicklungen zu leisten.

Zudem werfen MAS noch tiefer gehende Probleme auf. Agenten sind softwaretechnische Komponenten, die ausschließlich durch Algorithmen bestimmte berechenbare Funktionen ausführen, auch algorithmisch determinierte Verfahren maschinellen Lernens. Zielverfolgendes, kooperatives Verhalten kann ihnen mithin nur durch Programme – mittels definierter Nutzenfunktionen, Lernverfahren, Verhaltensrepertoires und geteilten »Ontologien« – vorgeschrieben werden. In der skizzierten abstrahierenden Perspektive »intentionaler Einstellung« werden MAS jedoch als quasi-soziale Systeme betrachtet und den Agenten Handlungsfähigkeit und Intentionalität zugeschrieben:

> »*An agent is an entity whose state is viewed as consisting of mental components such as beliefs, capabilities, choices, and commitments. These components are defined in a precise fashion, and stand in rough correspondence to their common sense counterparts. In this view, therefore, agenthood is in the mind of the programmer: What makes any hardware or software component an agent is*

> *precisely the fact that one has chosen to analyze and control it in these mental terms.«* (Shoham 1993, S. 52).

Verführt durch »Metaphernmigration« aus der Soziologie (»*social view of computing*«, ebd.), birgt diese anthropomorphisierende Betrachtungsweise die große Gefahr, dass das algorithmisch bestimmte adaptive Verhalten von Software-Agenten mit absichtsvollem Handeln von Menschen, dass die maschinelle Welt der Algorithmen und kontextfreien Daten mit der sozialen Welt von kontextabhängigen Bedeutungen, von Reflexion und Intentionalität menschlichen Handelns verwechselt wird. Und tatsächlich ist diese Begriffsverwirrung laufend zu beobachten, wenn etwa Daten mit Information gleichgesetzt werden oder wenn – wie in den Forschungsansätzen der »Sozionik« (Malsch 1998) oder der »Akteur-Netzwerk-Theorie« (ANT, Latour 1996) – technische Artefakte (z.B. Software-Agenten) und soziale Akteure als gleichartige, vernetzte »Aktanten« betrachtet werden, denen gleichermaßen Intentionalität und Handlungsfähigkeit zugeschrieben wird (Rammert 2003).

Mag diese Perspektive zur Untersuchung struktureller Verwandtschaften zwischen MAS und sozialen Systemen hilfreich sein, so ist sie für die Gestaltung soziotechnischer Systeme gänzlich unbrauchbar. In diesen Systemen sollen Menschen mit technischen Artefakten zweckmäßig und effizient zusammenwirken – was in den Vorstellungen zu Industrie 4.0 ausdrücklich vorgesehen ist. Dabei kommt es gerade entscheidend darauf an, den fundamentalen Unterschied zwischen algorithmisch bestimmtem, selbsttätigem Verhalten von artifiziellen Agenten und absichtsvollem, autonomem Handeln von Menschen zu beachten. Ein MAS mag zwar in seinem Verhalten als ein quasi-intentional agierender *»humunculus informaticus«* erscheinen, ist aber gleichwohl nur ein – wenn auch höchst problematisches – Arbeitsmittel in den Händen autonomer, kraft Sozialisation und Reflexionsfähigkeit verständig handelnder Menschen (Brödner 1997, S. 185 ff).

Eben darin wurzelt ein Kernproblem: Wie oben gezeigt, ist das Verhalten von MAS geschichtsabhängig, daher unter gegebenen Umständen nicht oder schwer zu verstehen und nicht vorhersehbar. Wie sollen Menschen sich aber solche Systeme aneignen, wie mit ihnen zweckmäßig und zielgerichtet interagieren, wenn diese sich in vergleichbaren Situationen jeweils anders und unerwartet verhalten? Das wäre ein eklatanter Verstoß gegen eine der Grundregeln der Mensch-Maschine-Interaktion, gegen die Forderung nach erwartungskonformem Verhalten. Zugleich würden auf Seiten der Nutzer stets aufs Neue überzogene Erwartungen an die »Handlungsfähigkeit« der Systeme geschürt. Konfrontiert mit diesen Widersprüchen, zugleich unter dem Erwartungsdruck erfolgreicher Bewältigung ihrer Aufgaben einerseits und angesichts des Verlusts der Kontrolle über Arbeitsmittel mit undurchschaubarem Verhalten andererseits, würden sie unter dauerhaften psychischen Belastungen zu leiden haben (so bereits Norman 1994).

Für die soziotechnische Systemgestaltung ist ferner von grundlegender Bedeutung, wie Aufgaben auf die automatisierten Systeme und die verbleibenden Menschen verteilt werden. Hinsichtlich dieses wesentlichen Gestaltungsaspekts wurden bereits früh anhand von qualifizierten Leitwarten-Tätigkeiten fundamentale »Ironien der Automatisierung« erkannt, die mit zunehmender Komplexität der technischen Systeme noch an Bedeutung gewinnen: Automatisierte Systeme wie MAS sollen möglichst weitreichend menschliche Arbeit ersetzen, deren Arbeitsvermögen aber im Störungs- oder Versagensfall der Systeme unersetzlich ist. Nun schwindet das menschliche Arbeitsvermögen aber dahin, je weniger es im automatisierten Normalbetrieb gebraucht wird. Dabei ist auf längere Sicht ein empfindlicher Verlust praktischer Handlungskompetenz zu verzeichnen, der aus ursprünglich kompetenten Nutzern am Ende hilflose, weil entwöhnte »Bediener« werden lässt (Bainbridge 1983; eindrückliche neuere Beispiele bei Carr 2013).

So nehmen denn in hoch automatisierten, eng gekoppelten Systemen wie MAS immer wieder »normale Katastrophen« (Perrow 1989) ihren Lauf:

- Der Totalverlust der Ariane V als einem fliegenden MAS beim Erstflug 1996 ist darauf zurückzuführen, dass der Überlauf einer Programmvariablen des Inertial-Lenksystems die Ausgabe von Statusdaten an den Navigationscomputer zur Folge hatte, die dieser als ›echte‹ Flugdaten verarbeitete. Aufgrund beträchtlicher ›Abweichungen‹ vom vorgesehenen Kurs wurden die Antriebsbooster bis zum Anschlag ausgelenkt mit der Folge, dass die Rakete sich wegen mechanischer Überbeanspruchung selbst zerstörte (Verlust 370 Mio. USD; http://de.wikipedia.org/wiki/Ariane_V88).
- Beim Hochfrequenzhandel an Börsen interagieren Handelscomputer wie ein MAS miteinander. Deren letztlich undurchschaubares Gesamtverhalten löst gelegentlich ungewollte Kursabstürze aus, so etwa beim Flash-Crash 2010 oder beim »Knightmare on Wall Street« der hierbei führenden Firma Knight, die 2012 binnen 45 min 440 Mio. USD verlor (heise online 4.8.2012).
- Der Absturz von Air France Flug AF 447 im Südatlantik 2009 ist laut abschließendem Untersuchungsbericht auf einen Strömungsabriss (*»stall«*) als Folge des Versagens der Geschwindigkeitsmesser und damit des Autopiloten zurückzuführen; die Piloten waren mit dieser Situation überfordert. Ähnliches gilt auch für einen weiteren »Stall«-Unfall eines Continental Connection Zubringer-Flugs in Buffalo 2009 (Carr 2013). Was wie hier stets als Pilotenfehler eingestuft wird, ist tatsächlich aber oft systematische Folge der »Automatisierungsironie« der mangels Übung verlorenen fliegerischen Kompetenz.

Darüber hinaus werfen Entwicklung und Gebrauch von MAS gewichtige, bislang freilich weitgehend ignorierte ethische Fragen auf: Dürfen Systeme mit derart undurchschaubarem Verhalten überhaupt von der Leine gelassen werden? Wie lässt sich dabei ein hinreichend sicherer Betrieb gewährleisten? Wer ist für allfälliges

Fehlverhalten und mögliche Schäden verantwortlich und haftbar zu machen? Sind es die Entwickler oder aber die Betreiber (oder gar die Benutzer) der Systeme? Und wie lässt sich das ggf. nachweisen? Wer haftet bei Schäden als Folge unglücklicher Verkettung äußerer Umstände, selbst wenn das System als solches funktioniert, wie es soll? Aktuell gewinnen diese Fragen im Zusammenhang von Experimenten mit selbstfahrenden Autos im öffentlichen Verkehr an Brisanz und es ist gut möglich, dass deren Zulassung letztlich an unzureichenden Antworten auf diese Fragen scheitert (Hilgendorf 2014).

Neben den heilsverkündenden Apologeten »künstlicher Intelligenz« melden sich zunehmend auch Apokalyptiker zu Wort, die diese Entwicklung als »größte Gefahr für die Menschheit« ansehen und ein Moratorium fordern (heise online 14.01.2015). Beiden ist gemeinsam, gegen Tatsachen weitgehend immun zu sein. So beruhen viele Konzeptionen »künstlicher Intelligenz«, so auch MAS, auf längst widerlegten philosophischen Grundlagen, auf die sie sich meist implizit stützen, gelegentlich gar explizit berufen. Darin ist der tiefere Grund zu sehen, dass Erwartungen wie Befürchtungen völlig überzogen sind:

- So erscheint der *Behaviorismus* (Pawlow 1927, Skinner 1953) zunächst als naheliegende theoretische Basis, da er das Verhalten von Lebewesen allein durch den Zusammenhang von außen beobachtbarer Reize und Reaktionen zu untersuchen und zu bestimmen trachtet. In dieser Hinsicht ist der Ansatz mit der Determination maschinellen Verhaltens vergleichbar. Dabei bleiben freilich die für das Handeln bedeutsamen mentalen Zustände, reflexives Bewusstsein und Intentionalität ausgeblendet.
- Als Ansatz zur Überwindung des Behaviorismus erkennt der *Funktionalismus* (Putnam 1960, Fodor 1968) mentale Zustände als wesentlich für die Erklärung von Verhalten an. Allerdings werden sie ungeachtet ihrer materiellen Realisierung als rein funktionale Zustände betrachtet nach dem Vorbild der Turing-Maschine. Freilich gilt längst auch diese Sichtweise als widerlegt, da – wie Putnam als ihr Begründer später (1991) durch scharfsinnige Gedankenexperimente gezeigt hat – gleiche funktionale Strukturen ganz unterschiedliche Weltbezüge (Gedanken und Erlebnisse) hervorzubringen vermögen.

In seiner neuesten Variante, einer in positivistischer Attitüde verengten Interpretation von »Embodiment« (vgl. u.a.Varela et al. 1991), werden CPS umstandslos mit lebendigen und einfühlsamen, zur Empathie und Reflexion ihres kontextbezogenen Erlebens und Handelns fähigen Körpern von Menschen gleichgesetzt (so zuletzt wieder Jeschke 2015). Übersehen wird dabei erneut die naturwissenschaftlich begründete Differenz deterministischen Verhaltens zu intentional gesteuertem, Sinn erzeugendem menschlichen Handeln und Verstehen im Kontext sozialer Praxis.

Tatsächlich beruhen aber ironischerweise die gelegentlich verblüffend »intelligent« erscheinenden Computerleistungen, entgegen den Postulaten ihrer Apolo-

geten, gar nicht auf spezifischen Konzepten der Forschung zur »künstlichen Intelligenz«, sondern nahezu allein auf exorbitant gesteigerter Rechenleistung. Damit wird ermöglicht, umfangreiche, einsichtsvoll und aufgabenangemessen implementierte Heuristiken mit schnellem Zugriff auf sehr große Bestände kodifizierten Wissens zu verarbeiten bzw. Lernverfahren bei sehr komplexen KNN durchzurechnen (sog. »Brute-force«-Methoden). Intelligent sind nicht die algorithmisch determinierten Computer, sondern deren Programmierer, die diese Möglichkeiten zur Bewältigung jeweils sehr spezifischer Aufgaben zu nutzen verstehen.

4.2 MAS und »holonische« Organisation von Produktionsprozessen

Im Zuge der Verwissenschaftlichung von Produktion und zunehmend wissensintensiver Wertschöpfung wird Wettbewerb weniger über Preise als über Qualität und vor allem Innovation ausgefochten. So sehen sich Organisationen gezwungen, in einem zunehmend dynamischen Umfeld voller Ungewissheit und geringer Planbarkeit zu agieren, was hohe Anforderungen an ihre Anpassungsfähigkeit stellt. Das ist der tiefere Grund dafür, dass ab den 1980er Jahren versucht wird, schwerfällige tayloristische Fertigungsstrukturen mit ihrer ausgeprägten horizontalen und vertikalen Arbeitsteilung und ihrer zentralen Steuerung mittels Planung, Weisung und Kontrolle durch wandlungsfähigere, objektorientierte zellulare Strukturen mit weitgehend autonomen und auftragsgebunden selbstgesteuerten Einheiten geringer Arbeitsteilung zu ersetzen, die höhere Produktivität und weit geringere Durchlaufzeiten aufweisen und sich für agile Produktion leichter koordinieren lassen (Brödner 2008, 2010).

Systematische Anwendung dieser Strukturierungsprinzipien führt – angelehnt an eine Begriffsbildung von Koestler (1967) – zu sog. »holonischen Organisationen« als besonders entwicklungs- und anpassungsfähigen Systemen. Sie gliedern sich rekursiv über mehrere Ebenen in relativ autonome Teilsysteme (»Holone«). Diese sind weder Teile im Sinne bloßer Komponenten des Ganzen noch völlig unabhängige Ganzheiten. Vielmehr handelt es sich um weitgehend unabhängig operierende, sich selbst steuernde und anpassende Einheiten, die sich zu einem übergeordneten Ganzen fügen, indem sie ihre Operationen koordinieren. Jedes Holon (Organisationszelle) verfügt über die nötigen Möglichkeiten der Anpassung, reflexiven Fähigkeiten und Mittel, seine operativen Aufgaben im Rahmen des Ganzen zu erledigen und sich veränderlichen Bedingungen anzupassen. Dazu besitzt es eine Aufgabenbeschreibung (Modell) und beobachtbare operative Ziele, es vermag relevante Veränderungen ihrer Umwelt wahrzunehmen und seine operativen Prozesse durch Lernen zielorientiert anzupassen im Sinne gestalterhaltender Selbstregulation und produktiven wie innovativen Umgangs mit Umweltveränderungen. Das jeweils übergeordnete System sorgt für Integration und Kohärenz, indem es untergeordnete operative Ziele und Aufgaben koordiniert, deren Erfüllung überwacht und die notwendigen Ressourcen bereitstellt (Koestler 1967, Mathews 1996).

Im vorliegenden Zusammenhang sind holonische Organisationsformen bedeutsam, weil sich in den letzten beiden Dekaden im Rahmen des internationalen Forschungsprogramms »Intelligent Manufacturing Systems (IMS)« und auch entsprechender EU-Förderung die Entwicklung von MAS für die Produktion stark an holonischen Organisationsprinzipien orientiert hat (vgl. Abb. 2 links). Dazu gibt es eine Fülle von – freilich hauptsächlich akademischer – Literatur (zur Übersicht vgl. Farid 2004, ein typisches Beispiel in Peschl et al. 2011). Allerdings wird dabei mit der Einschränkung auf algorithmisch determiniertes Verhalten von MAS erneut ein Reduktionismus betrieben, gegen den Koestler mit dem Holon-Konzept einst ausdrücklich zu Felde zog.

So scheint paradoxerweise die Realisierung flexibel automatisierter Produktionsprozesse mittels MAS bzw. »cyber-physischer Systeme« deren durchgängige Restrukturierung nach holonischen Organisationsprinzipien vorauszusetzen, um aussichtsreich zu funktionieren. Allerdings wurde holonische Reorganisation (im umfassenden Sinn von Koestler, vgl. Mathews 1996) bereits seit den 1980er Jahren beim Übergang zu wissensintensiver Wertschöpfung systematisch betrieben (etwa in Gestalt von integrierten Konstruktionsteams und Fertigungszellen); diese Bemühungen verstrickten sich freilich oft in den Schwierigkeiten organisationalen Wandels. In real existierenden Organisationen ist dieser mit tiefgreifenden und schwierig zu bewältigenden Veränderungen von Denkweisen, Handlungsroutinen und Machtbeziehungen verbunden, an denen bisher viele Organisationen scheitern. Allein schon aus diesem Grund ist ein Erfolg der MAS-basierten Entwicklungsstrategie fraglich.

4.3 Probleme im Umgang mit Big Data

Auch der Umgang mit Big Data wirft gleich eine ganze Reihe schwerwiegender, teils unlösbarer Probleme auf. Neulich hat in typischer Manier technikeuphorischen Überschwangs der Chefredakteur der Internet-Zeitschrift »Wired« vollmundig »das Ende von Theorie« verkündet: Große Datenmengen könnten theoretisch angeleitete Forschung ablösen, allein auf Korrelationen beruhende Vorhersagen seien Hypothesen-basierten Prognosen überlegen und Korrelation ersetze Kausalität (Anderson 2008). Hinter dieser unfasslichen, aber verbreiteten und verführerischen Narretei steht der alt bekannte Trugschluss »*cum hoc ergo propter hoc«:* Wenn zwei Ereignisse a und b zusammen auftreten, kann man ohne aufwendige zusätzliche Analysen niemals wissen, ob das Ereignis a durch b oder umgekehrt b durch a hervorgerufen wurde, man kann auch nicht wissen, ob beide mit einem gemeinsamen dritten, unerkannten Ereignis zusammenhängen oder ob sie rein zufällig zusammen auftreten. Im Extrem führt das zu dem apophenischen Wahn, in Haufen sinnloser Daten Muster zu erkennen.

Zudem werden kontext- und sinnfreie Daten ständig mit bedeutungsvoller Information aus kontextabhängiger Interpretation verwechselt, wobei suggeriert wird, dass Daten allein schon vermeintlich Information, mithin Bedeutung und

Geltung zukomme. Zwar lassen sich bei Kenntnis der Semantik der entsprechenden Verarbeitungsalgorithmen durch gegenstandsbezogene Kombination verschiedener Daten (z.B. in Personenprofilen) Bruchstücke von Kontext rekonstruieren und damit der mögliche Interpretationsrahmen einengen, der gleichwohl unvollständig rekonstruierte Kontext lässt aber immer noch ganz verschiedene Interpretationen und damit auch Fehlinterpretationen zu (Fälle fehlinterpretierter Schufa-Daten etwa legen davon eindrücklich Zeugnis ab). Zudem beschreiben Daten oft eine Realität, die durch die Beschreibung erst geformt wird: Aus deskriptiven können normative Daten, aus Häufigkeiten Gewissheiten gemacht oder interessengeleiteten Deutungen kann machtvoll Geltung verschafft werden (Boyd 2011). Mit Daten wird so Objektivität und Faktizität suggeriert, wo die tatsächliche Bedeutung sich erst aus dem vollständigen (aber nicht verfügbaren) Kontext ergäbe. Aus eben diesen Gründen gelten vor Gericht Indizienbeweise als höchst problematisch.

Außerdem ist bei großen Datenbeständen, die auch noch aus unterschiedlichen Quellen stammen, die Datenqualität mangelhaft: Daten sind oft nicht repräsentativ, fehlerhaft, obsolet oder gar inkonsistent. Auch lässt sich das Ausmaß dieser Mängel oft nicht einmal abschätzen oder beurteilen. Wenn der Umgang mit Big Data aber nicht anerkannten, streng methodischen Regeln statistischen Schließens folgt, die auch gesichertes Wissen über die Datenqualität einschließen, muss er als pseudowissenschaftlich verbrämtes Kaffeesatzlesen angesehen werden. Wenn Daten zum »bedeutendsten Wirtschaftsgut«, zum »Rohöl des 21. Jahrhunderts« erklärt werden, bedürfen sie wie dieses sorgfältiger und aufwendiger Raffinerie, um valide Information daraus zu gewinnen.

Schließlich bereitet die mangelnde Datensicherheit große Probleme. Erhebliche Risiken ergeben sich für Unternehmen durch Datenverlust und Datendiebstahl, durch Spionage von außen oder Sabotage von innen (wovon derzeit fast tägliche Berichte über »Cyber- Angriffe« zeugen). Diese Risiken steigern sich noch, wenn, wie zunehmend üblich, Daten und Verarbeitungsprozesse an Service Provider oder in die »Cloud« ausgelagert werden. Angesichts von Häufigkeit und Schadensvolumen ist es nachgerade unverständlich, warum Unternehmen ihre im Betrieb von Industrie 4.0 laufend anfallenden, detailreichen, produkt- wie prozessbezogenen, mithin höchst wettbewerbssensiblen Datenströme diesen Risiken aussetzen. Zwar können und müssen laufend technische und organisatorische Sicherheitsvorkehrungen getroffen werden, diese schützen aber niemals hinreichend, denn jede noch so ausgeklügelte Schutzmaßnahme kann, wie die Praxis lehrt, auch wieder überwunden werden.

5 Zum Schluss: Perspektiven künftiger Entwicklung

Für die Gestaltung soziotechnischer Systeme, insbesondere für Entwicklung, Einsatz und Gebrauch von Computersystemen in der Produktion, sind schon seit jeher zwei entgegengesetzte Perspektiven im Spiel:

- Die technikzentrierte Perspektive weitestgehender Automatisierung von Wissensarbeit wie sie in den Bestrebungen zur »künstlichen Intelligenz« – AI (Artificial Intelligence) – angelegt ist: »Smart machines« und »autonome Agenten«, zu MAS vernetzt und mit Big- Data-Methoden kombiniert, sollen menschliches Arbeitsvermögen in der Produktion nachahmen und weitgehend ersetzen; deren »Lernfähigkeit« – tatsächlich nur algorithmisch gesteuerte Adaption an Umweltgegebenheiten – soll gleichwohl hinreichende Flexibilität der Anpassung an wechselnde Anforderungen gewährleisten (gemäß der »intentional stance«; Minsky 1988, Shoham 1993, Wooldridge 2002).
- In der praxistheoretischen Perspektive menschengerecht und aufgabenangemessen gestalteter, als Werkzeug und Kooperationsmedium angeeigneter und genutzter Computersysteme – IA (Intelligence Amplification) – sollen diese lebendige Arbeit auf eine Weise unterstützen, dass die Entfaltung von deren Arbeitsvermögen, mithin die Steigerung ihrer Produktivkraft und Innovationsfähigkeit, ermöglicht und gefördert wird: »Things that make us smart« (Norman 1993; vgl. auch Ehn 1988, Winograd 1996).

Die technikzentrierte Perspektive erscheint gerade auch aufgrund kümmerlicher historischer Erfahrungen mit Konzepten »künstlicher Intelligenz«, etwa mit Expertensystemen und CIM, als wenig erfolgversprechend, eher als Verschwendung von Ressourcen. Bei evidenzbasierter Betrachtung beruht der säkulare Erfolg des Einsatzes und Gebrauchs von Computersystemen stattdessen ganz überwiegend auf der praxistheoretisch angeleiteten Perspektive der »Intelligenzverstärkung« und den damit verbundenen Methoden der Organisationsentwicklung. Dabei werden menschliche Reflexions- und Lernfähigkeit mit maschineller Präzision und Geschwindigkeit verknüpft. Ihr muss daher alle Aufmerksamkeit gelten, um Flexibilität mit Effizienz zu verbinden. Dabei muss sich die soziotechnische Gestaltung ganz an den Eigenheiten und Bedürfnissen menschlichen Handelns und sozialer Praxis sowie an den Bedingungen der Entfaltung praktischen Arbeitsvermögens orientieren, um Produktivität und Kreativität zu ermöglichen. Gefordert sind dauerhaft kompetenzerhaltende und lernförderliche Arbeitsaufgaben, durchschau- und beherrschbare, aufgabenangemessene Arbeitsmittel mit erwartungskonformem Verhalten sowie ausreichende Zeitressourcen zur Aneignung der Arbeitsmittel und zu laufender Optimierung von Prozessen (Brödner 2008).

Der reflexiven Dynamik der Explikation von Können als Wissen und der Aneignung von Wissen als erweitertem Können zufolge wird mit dem Einsatz von Computersystemen in Organisationen massiv in deren soziale Praktiken interveniert, oft mit überraschenden Folgen. So erweist sich etwa praktisches Können durch den Gebrauch angemessen gestalteter Artefakte wiederholt »intelligenten«,

die menschliche Expertise ersetzenden Automaten als überlegen, sogar dann noch, wenn die Automaten leistungsfähiger als menschliche Experten sind. Beispielsweise hat der Schachweltmeister Kasparov ein zunächst seinem Können überlegenes Computersystem (von vergleichbarer Leistung wie »Deep Blue« von IBM, dem er unterlag) wiederum geschlagen, indem er seinerseits einen wesentlich einfacher gestalteten Computer als Werkzeug zu Hilfe nahm (Kasparov 2010).

Nach diesem Muster des Zusammenwirkens von Menschen und Computer-Artefakten können beispielsweise an CPS laufend erfasste Daten als Grundlage für interaktive Assistenzsysteme mit gebrauchstauglich gestalteter Benutzungsoberfläche zur Rekonfigurierung oder Optimierung von Prozessen, zu deren wirksamer Simulation und Steuerung oder auch zur datengestützten Diagnose von CPS genutzt werden. Dabei ist zwecks gelingender Interaktion wichtig, Benutzern die Möglichkeit zu bieten, in steuerbaren Detaillierungsgraden Einblick in den Verlauf von Maschinenzuständen, in gegebene Einstellungen oder verwendete Methoden zu nehmen. Nur so können sie sich, um zweckmäßig einzugreifen, ein eigenes Bild von den maschinellen Abläufen und dem Zustandekommen von Resultaten machen.

Dieser praxistheoretisch begründeten »IA«-Perspektive zu folgen, hieße, statt auf illusionäre »AI«-Hoffnungen zu setzen, höhere Flexibilität, Produktivität und Innovationsfähigkeit weit wirksamer durch soziotechnische Gestaltung »guter Arbeit« zu erreichen, auf Basis reflexiven und kreativen Zusammenwirkens kompetenter Experten mit gebrauchstauglich gestalteten Computersystemen und dadurch ermöglichter Entfaltung lebendigen Arbeitsvermögens.

6 Literatur

Anderson, C. 2008: The End of Theory, Wired 23.06.08

Bainbridge, L. 1983: Ironies of Automation, Automatica 19, S. 775-779

Barney, J. B. 1991: Firm Resources and Sustained Competitive Advantage, Journal of Management 17 (1), S. 99–120

Bell, D. 1975: Die nachindustrielle Gesellschaft, Frankfurt/M.

BMBF 2014: Innovationen für die Produktion, Dienstleistung und Arbeit von morgen, Bonn

Boyd, D. 2011: Six Provocations for Big Data, www.softwarestudies.com/cultural_analytics/Six_Provocations_for_Big_Data.pdf [zuletzt aufgesucht am 10.02.2015]

Breadshaw, J.M. 1997: An Introduction to Software Agents. In: Breadshaw, J.M. & Hutchinson, F. (eds.): Software Agents, Cambridge (MA), S. 3-46

Brödner, P., 2010: Wissensteilung und Wissenstransformation. In: Moldaschl, M. & Stehr, N. (Hg.): Wissensökonomie und Innovation. Beiträge zur Ökonomie der Wissensgesellschaft, Marburg, S. 455-480

Brödner, P. 2008: Das Elend computerunterstützter Organisationen. In: Gumm, D. et al. (Hg.): Mensch –Technik – Ärger? Zur Beherrschbarkeit soziotechnischer Dynamik aus transdisziplinärer Sicht, Münster, S. 39-60

Brödner, P. 2006: Betriebliche Rationalisierungsstrategien und Einsatz technischer Systeme. In: Zimolong, B.;

↑Inhalt↑

Konradt, U. (Hg.): Ingenieurpsychologie. Enzyklopädie der Psychologie: Wirtschafts-, Organisations- und Arbeitspsychologie - Band 2, Göttingen, S. 943-980

Brödner, P. 1997: Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen, Berlin

Broy, M. (Hg.) 2010: Cyber-physical systems. Innovation durch softwareintensive eingebettete Systeme, Berlin Heidelberg: Springer

Brynjolfsson, E. & McAfee, A. 2014: The Second Machine Age. Wie die nächste digitale Revolution unser aller Leben verändern wird, Kulmbach

Carr, N. 2013: All Can Be Lost: The Risk of Putting Our Knowledge in the Hands of Machines, The Atlantic No. 11

Cyranek, G. & Ulich, E. (Hg.) 1993: CIM – Herausforderung an Mensch, Technik, Organisation. Stuttgart: Schäffer Poeschel und Zürich: vdf

Dedrick, J.; Gurbaxani, V. & Kraemer, K. L. 2003: Information Technology and Economic Performance: A Critical Review of the Empirical Evidence, ACM Computing Surveys 35, S. 1-28

Dennett, D. C. 1987: The Intentional Stance, Cambridge (MA)

Drucker, P. F. 1994: The Age of Social Transformation, The Atlantic No. 11, S. 53-80

Ehn, P. 1988: Work-Oriented Design of Computer Artifacts, Stockholm

Farid, A. M: 2004: A Review of Holonic Manufacturing Systems Literature, University of Cambridge

Fodor, J. 1968: Psychological Explanation, New York

Foerster, H.v. 1993: Wissen und Gewissen, Frankfurt/M.

Forschungsunion & acatech 2013: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0, Frankfurt/M.

Giddens, A. 1988: Die Konstitution der Gesellschaft, Frankfurt/M.

Grant, R.M. 1996: Toward a Knowledge-based Theory of the Firm, Strategic Management Journal 17, S. 109-122

Hilgendorf, E. 2014: Recht, Maschinen und die Idee des Posthumanen, Telepolis 24.05.2014

Hunt, V. D. 1989: Computer Integrated Manufacturing Handbook. London New York: Kluwer Academic Publishers

Jeschke, S. 2015: Auf dem Weg zu einer »neuen KI«: Verteilte intelligente Systeme, Informatik Spektrum 38 (1), S. 4-9

Jorgenson, D. W.; Ho, M. S. & Stiroh, K. J. 2008: A Retrospective Look at the U.S. Productivity Growth Resurgence, Journal of Economic Perspectives 22 (1), S. 3-24

Kasparov, G. 2010: The Chess Master and the Computer, The New York Review of Books 11.02.2010

Koestler, A. 1967: The Ghost in the Machine, London

Kriesel, D. o.J.: Kleiner Überblick über Neuronale Netze, http://www.dkriesel.com/science/neural_networks [zuletzt aufgesucht am 12.02.2017]

Latour, B. 1996: On Actor-Network Theory. A Few Clarifications, Soziale Welt 47, S. 369-381

Maes, P., 1994: Agents that Reduce Work and Information Overload, CACM 37 (7), S. 31-41

Malsch, T. (Hg.): 1998: Sozionik. Soziologische Ansichten über künstliche Sozialität, Berlin

Mathews, J. A. 1996: Holonic Organisational Architectures, Human Systems Management 15 (1), S. 27-54

Minsky, M. 1988: The Society of Mind, New York

Norman, D. A. 1994: How Might People Interact with Agents, CACM 37 (7), S. 68-71

Norman, D. A. 1993: Things that Make Us Smart, Reading (MA)

Penrose, E. T. 1995: The Theory of the Growth of the Firm, 3rd edn., New York Oxford

Perrow, C. 1989: Normale Katastrophen, Frankfurt/M.

Peschl, M.; Link, N.; Hoffmeister, M.; Gonçalves, G.& Almeida, F. L. F. 2011: Design and Implementation of an Intelligent Manufacturing System, JIEM 4 (4), S. 718-145

Peters, K. & Sauer, D. 2006: Epochenbruch und Herrschaft. Indirekte Steuerung und die Dialektik des Übergangs. In: Scholz, D. et al. (Hg.): Turnaround? Strategien für eine neue Politik der Arbeit, Münster, S. 98-125

Prasse, M. & Rittgen, P. 1998: Bemerkungen zu Peter Wegners Ausführungen über Interaktion und Berechenbarkeit, Informatik-Spektrum 21, S. 141-146

Putnam, H. 1960: Minds and Machines. In: Hook, S. (ed.): Dimensions of Mind, New York

Putnam, H. 1991: Representation and Reality, Cambridge (MA)

Rammert, W. 2003: Technik in Aktion. In: Christaller, T. & Wehner, J. (Hg.): Autonome Maschinen – Perspektiven einer neuen Technikgeneration, Wiesbaden, S. 289-315

Reckwitz, A., 2003: Grundelemente einer Theorie sozialer Praktiken. Eine sozialtheoretische Perspektive, Zeitschrift für Soziologie 32 (4), S. 282-301

Schmidhuber, J. 2015: Deep Lerning in Neural Networks. An Overview, Neural Networks 61, S. 85–117

Sharif, M. et al. 2016: Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition, ACM Conference on Computer and Communication Security, Vienna

Shoham, Y. 1993: Agent-Oriented Programming, Artificial Intelligence 60, S. 51-92

Skinner, B. F. 1953: Science and Human Behavior. New York

Varela, F. J.; Thompson, E. & Rosch, E., 1991: The Embodied Mind. Cognitive Science and Human Experience. Cambridge (MA)

Wick, C. 2017: Deep Learning, Informatik Spektrum 40 (1), S. 103-107

WEF 2012: The Future of Manufacturing. Opportunities to drive economic growth, a World Economic Forum Report in collaboration with Deloitte Touche Tohmatsu Limited

Wegner, P. 1997: Why Interaction is More Powerful than Algorithms, CACM 40 (5), S. 80-91

Wiener, N. 1950/66: Mensch und Menschmaschine. Kybernetik und Gesellschaft, Frankfurt/M: Athenaeum (engl. Original 1950: The Human Use of Human Beings. Cybernetics and Society, Boston (MA))

Winograd, T. 1996: Bringing Design to Software, Reading (MA)

Wooldridge, M. 2002: An Introduction to Multi-Agent Systems, New York

↑Inhalt↑

Grenzen und Widersprüche der Entwicklung und Anwendung ›Autonomer Systeme‹

1 Einführung: Überzogene Erwartungen

Derzeit überrollt uns nach drei Jahrzehnten relativer Stille erneut eine Sturzflut von Meldungen über Projekte vermeintlicher ›künstlicher Intelligenz‹ (KI, engl. AI für ›artificial intelligence‹). Die Spannweite der Meldungen reicht von Heilsversprechen über Erfolgsgeschichten bis zu Szenarien der Apokalypse. Als kleine Auswahl aus dem gesamten Spektrum vermitteln die nachstehenden Äußerungen einen Eindruck:

- Unter der Überschrift »The Future AI Company« wird die künftige Automobilfabrik so beschrieben: »Superschlaue Computer, die ständig lernen, werden vieles übernehmen, was bisher Menschen erledigen: Sie antworten, wenn Kunden oder Lieferanten fragen, automatisch […] [S]ie entwerfen sogar Autos und rechnen aus, wie sich die Entwürfe in der Fabrik umsetzen lassen.« (Hofmann, VW-Vorstand, laut Schäfer 2016),
- »Lernende Maschinen erkennen Gesichter und bringen sich selbst das Schachspielen bei.« (Dworschak 2018),
- »Invasion der Roboter: Künstliche Intelligenz ist bald so normal wie Strom.« (Recke 2016),
- Laut Google-CEO Pichai ist künstliche Intelligenz für die Menschheit bedeutender als die Entdeckung des Feuers oder die Entwicklung der Elektrizität (Bastian 2018),
- Hawking fears »AI may replace humans altogether« as a »new form of life that will outperform humans.« (Sulleyman 2017).

Gegenüber solchen offenkundig sensationslüsternen, auf Verblüffung der Öffentlichkeit angelegten Äußerungen gilt es zunächst einmal den gesunden Menschenverstand zu bewahren und Aufklärung in wissenschaftlicher Analyse zu suchen. Auffällig an diesen und ähnlichen Äußerungen ist zunächst eine Gemeinsamkeit: Heilsverkünder/-innen wie Apokalyptiker/-innen sind gleichermaßen gegen Tatsachen immun. Gegen darin zum Ausdruck kommende ›KI‹-Wahnvorstellungen hilft daher nur, die Wahrheit in relevanten Tatsachen zu suchen. Die Probleme beginnen freilich schon damit, dass es bis heute nicht gelingt, ›KI‹-Systeme logisch

zufriedenstellend von ›gewöhnlichen‹ Computersystemen zu unterscheiden. Zwecks »Maschinisierung von Kopfarbeit« (Nake 1992) in Zeichenprozessen sozialer Praxis werden beide, auch gewöhnliche Computersysteme, immer schon zur Bewältigung von Aufgaben geschaffen »commonly thought to require intelligence« (so eine übliche ›KI‹-Definition, vgl. Autorengruppe 2018). Dazu führen beide Systemarten gleichermaßen berechenbare Funktionen zur planvollen automatischen Verarbeitung zugehöriger Daten aus (vgl. Brödner 2008).

Bemerkenswert an Äußerungen wie den zitierten ist ferner, dass sie allesamt bestimmte Computerartefakte als solche in den Blick nehmen und diesen die Verblüffung hervorrufenden Eigenschaften zuschreiben. Dieser Fokussierung auf technische Artefakte liegt aber ein ebenso verbreitetes wie tief gehendes Missverständnis von Technik im Allgemeinen und der Computertechnik im Besonderen zugrunde. Einem Bonmot des Philosophen Ortega y Gasset (1949) zufolge ist Technik »die Anstrengung, Anstrengungen zu ersparen«. Damit trifft er den Kern der Sache: Genauere Analyse zeigt nämlich, dass Technik, verstanden als bloße Ansammlung technischer Artefakte, viel zu kurz greift. Artefakte fallen nicht vom Himmel, sondern müssen für bestimmte Zwecke mühsam konzeptionell entwickelt und materiell hergestellt werden. Als solche sind sie aber bloß tote, nutzlose Gegenstände, solange sie nicht für bestimmte Aufgaben zweckgemäß eingesetzt, mithin dafür angeeignet und praktisch wirksam verwendet werden. Das alles geschieht im Spannungsfeld des technisch Machbaren, der Formbarkeit von Natur, und des sozial Wünschenswerten, abhängig von jeweils herrschenden Interessen.

Nach allgemeinem professionellem Verständnis wird Technik daher definiert als die Gesamtheit von Maßnahmen zur Herstellung und zum Gebrauch künstlicher Mittel für gesellschaftliche Zwecke. Ihr werden damit nicht nur die Artefakte und Sachsysteme selbst zugerechnet, sondern gerade auch deren sozial konstruierte und kulturell vermittelte Herstellung und Anwendung (vgl. Ropohl 1991, VDI 1991). Als geronnene Erfahrung verkörpern sie ein Stück sozialer Praxis und als Arbeitsmittel stellen sie Handlungsanforderungen an ihren Gebrauch, durch den Artefakte erst ihren Sinn erhalten und in ihrer Qualität zu beurteilen sind. Eben in den Prozessen der Entwicklung und Herstellung technischer Artefakte sowie ihrer Aneignung zu praktisch wirksamer Verwendung liegen gerade die eigentlichen Probleme der »Anstrengung, Anstrengungen zu ersparen«; eben hierin liegen auch die Wurzeln missbräuchlichen Umgangs.

Sich in dieser Perspektive einigen der Probleme fortgeschrittener Computerentwicklung zu stellen, ist das Anliegen des vorliegenden Beitrags. Dazu werden im Folgenden zunächst am Beispiel sogenannter ›künstlicher neuronaler Netze‹ (KNN) und Verfahren des ›Deep Learning‹, die derzeit als vermeintliche Schlüsseltechnik der ›KI‹ besonders hoch im Kurs stehen, deren Entwicklungsprobleme und Funktionsweisen aufgezeigt und hinsichtlich ihrer Tragweite beurteilt. Sodann werden die besonderen Einsatz- und Anwendungsprobleme beleuchtet, die mit diesen Systemen im Vergleich zu herkömmlichen Computersystemen

verbunden sind. Vor dem Hintergrund eines kurzen historischen Exkurses über Meilensteine der Entwicklung von Computertechnik und Computing Science wird anschließend über verbreitete, tief sitzende Missverständnisse von Funktionsweise und Leistungsgrenzen von Computern aufgeklärt. Darauf fußend wird eine abschließende Bewertung vorgenommen.

2 ›Deep Learning‹ und seine unterschätzten Entwicklungsprobleme

Neben Verfahren zur Analyse von ›Big Data‹ auf Basis von Methoden schließender Statistik bieten sogenannte ›künstliche neuronale Netze‹ (KNN) einen zweiten grundlegenden Ansatz der Verwirklichung sog. ›maschinellen Lernens‹. Infolge exponentiell gesteigerter Leistung der Computer-Hardware sind sie in letzter Zeit zu einem bevorzugten Gegenstand der ›KI‹-Forschung geworden. In Form vielschichtig strukturierter adaptiver Netzwerke bieten sie Verfahren zum sog.›Deep Learning‹ (LeCun et al. 2015, Schmidhuber 2015) als einer Art neuer ›Wunderwaffe‹.

Im Unterschied zu früheren ›KI‹-Ansätzen der ›allgemeinen Problemlösung‹ (Simon & Newell 1972) oder der ›wissensbasierten Systeme‹ der ›symbolischen KI‹ bzw. des ›Cognitive Computing‹ (IBM), wie sie u.a. beim ›Computer Integrated Manufacturing‹ der 1980er Jahre verfolgt wurden, setzt die heute dominante neue Welle der ›KI‹-Forschung nach einer langen Periode stark gebremster Aktivität nun vornehmlich auf die Entwicklung und den Einsatz von KNN als einem Grundmodell ›lernfähiger‹ Systeme. Dieser Ansatz ist freilich keineswegs neu, sondern geht in seinen Anfängen zurück auf das »Perceptron« (McCulloch/ Pitts 1943) als einer biologisch inspirierten Nachahmung der logischen Funktionsweise von Nervenzellen bereits in der Zeit der allerersten materiellen Realisierungen von Computern. In Gestalt der »Lernmatrix« von Steinbuch 1961 wieder aufgegriffen und zu größeren Netzwerken verknüpft (vgl. Hilberg 1995), fiel dieser Ansatz nach einem Verdikt durch Minsky und Papert (1969) erneut in einen Dornröschenschlaf, aus dem er in den 1990er Jahren langsam wieder erwachte. Wachgeküsst wurde er von der Aussicht auf ›maschinelles Lernen‹ im Zusammenhang mit der Einsicht, dass die sich auf die logische Verarbeitung von durch Daten repräsentiertem explizitem Wissen stützende ›symbolische KI‹ ihrerseits an den Grenzen der Explizierbarkeit von Können bzw. implizitem Wissen gescheitert war (vgl. Brödner 1997).

Versuche, kognitive Leistungen des Menschen, insbesondere sein intelligentes Handeln nachzuahmen, legen die Grundidee und den Ansatz nahe, das Gehirn mit seinen vielschichtig vernetzten Neuronen als Vorbild zu nehmen und einige seiner Strukturmerkmale und Funktionen möglichst direkt in Computersystemen nachzubilden. Freilich dürfen solche konnektionistischen Modelle nicht missverstanden werden als ›naturgetreue‹ Nachbildungen des Gehirns oder des Zentralner-

vensystems, allenfalls gibt es gewisse strukturelle und funktionale Ähnlichkeiten, die durch den Aufbau des Gehirns inspiriert werden.

Ein KNN besteht im Wesentlichen aus einer Menge miteinander verbundener *Knoten*, die in Abhängigkeit von ihrem aktuellen Aktivierungszustand und der momentanen Eingabe ihren neuen Zustand bestimmen und eine Ausgabe produzieren. Diese Elemente sind gemäß der *Netzwerkstruktur* miteinander verknüpft. Diese kann als gerichteter gewichteter Graph oder durch eine *Konnektionsmatrix* dargestellt werden. Die Dynamik eines KNN wird beschrieben durch (vgl. Abb. 1):

- eine *Propagierungs-* bzw. *Übertragungsfunktion* net_j, die aus den Ausgaben der vorgeschalteten Elemente sowie der Gewichtung der Verbindungen die aktuellen Eingaben in interne Netzwerkelemente berechnet,
- eine *Aktivierungsfunktion* σ, die für jedes Element dessen Aktivierung *oj* als Ausgabe bestimmt abhängig davon, ob ein Schwellwert von der Netzeingabe *netj* überschritten wird oder nicht.

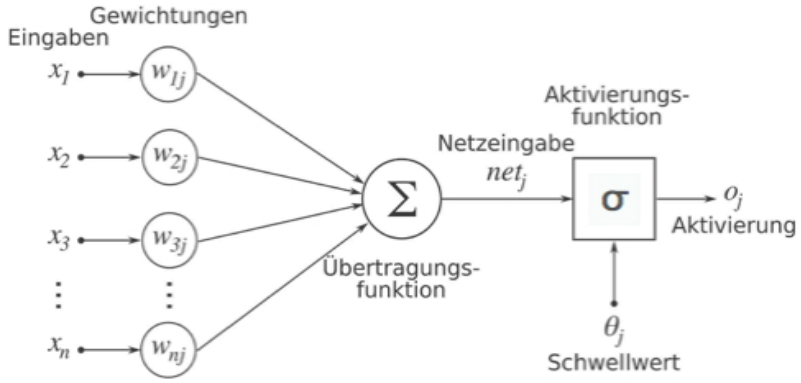


Abb. 1: *Berechnungsfunktionen an einem Netzknoten (Quelle: Wikipedia CC BY-SA 3.0)*

In der Regel besteht die Propagierungsfunktion aus einer einfachen Summenbildung der gewichteten Verbindungseinflüsse und die Aktivierungsfunktion wird meist für alle Elemente des Netzwerks einheitlich festgelegt. Innerhalb eines Netzwerks wird noch zwischen Eingabe-, Ausgabe- und internen Elementen unterschieden (›hidden units‹, die zu tief gestaffelten ›hidden layers‹ zusammengefasst werden, daher die Bezeichnung ›Deep Learning‹). Mittels verschiedener Typen von Ausgabe-, Propagierungs- und Aktivierungsfunktionen können Klassen von konnektionistischen Modellen gebildet werden.

Um ihre Aufgaben zu lösen, müssen KNN während des Entwurfs passend strukturiert und ihre Prozessoren mit einem bestimmten ›Lern‹-Algorithmus gesteuert werden. Das Netz als Ganzes wird über diese Festlegungen hinaus nicht programmiert, sondern passt sich durch Veränderung der Gewichte nach Maßgabe eines den Nutzen maximierenden ›Lern‹-Algorithmus an die spezielle Aufgabenstellung an (daher die Benennung *adaptiv*). Beispielsweise werden bei Problemen

der Musteridentifikation oder der Klassifikation – ein Aufgabentyp, bei dem künstliche neuronale Netze, vor allem solche vom Typ der ›faltenden‹ oder ›Convolutional Neural Networks (CNN)‹, besonders leistungsfähig sind – einer langen Reihe von Eingabemustern jeweils die zugehörigen Klassen als Ausgänge zugeordnet; aus diesen Zuordnungen vermag dann der Algorithmus mittels einer Nutzenfunktion automatisch passende Verbindungsgewichte zu bestimmen. Dies funktioniert auch bei Mustern, die durch explizite Merkmalsbeschreibungen schwierig oder gar nicht zu fassen sind (etwa bei der Identifikation handgeschriebener Buchstaben) – freilich mit Unsicherheiten. Meist ist dafür allerdings eine sehr große Zahl von Trainingsbeispielen (in der Größenordnung von 10^6) erforderlich (vgl. LeCun et al. 1998).

Bei der Anpassung ist für die Bestimmung der Gewichte $W_j := W_j - \eta \nabla_W L$ eines KNN die Berechnung des Gradienten $\nabla_W L(N(x))$ einer Nutzen- oder Verlustfunktion $L(N(x))$ erforderlich (mit $N(x)$ als Netzwerksausgabe, $L(N)$ als etwa über alle Trainingsbeispiele summierter euklidischer Distanz sowie der »Lernrate« η). Mit der Aktivierungsfunktion $\sigma(x)$ lässt sich – am einfachen Beispiel eines dreischichtigen KNN – die Funktionsweise des Backpropagation-Algorithmus und das häufig auftretende Problem des »schwindenden Gradienten« aufzeigen:

$$N(x) = W_1 \cdot \sigma(\overbrace{W_2 \cdot \sigma(\underbrace{W_3 \cdot x}_{N_3})}^{N_2})$$

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

$$\begin{aligned}
\frac{\mathrm{d}L}{\mathrm{d}W_1} &= \sigma(N_2) \cdot \frac{\mathrm{d}L}{\mathrm{d}N} \\
\frac{\mathrm{d}L}{\mathrm{d}W_2} &= \sigma(N_3) \frac{\mathrm{d}\sigma(N_2)}{\mathrm{d}N_2} \cdot W_1 \cdot \frac{\mathrm{d}L}{\mathrm{d}N} \\
\frac{\mathrm{d}L}{\mathrm{d}W_3} &= x \cdot \frac{\mathrm{d}\sigma(N_3)}{\mathrm{d}N_3} \cdot W_2 \cdot \frac{\mathrm{d}\sigma(N_2)}{\mathrm{d}N_2} \cdot W_1 \cdot \frac{\mathrm{d}L}{\mathrm{d}N}
\end{aligned}$$

Dabei wiederholt auftretende Faktoren sind während des Trainings oft < 1, so dass das Ergebnis infolge ihrer Multiplikation insgesamt gegen Null tendiert – daher der »schwindende Gradient« und die Schwierigkeit, vielschichtige KNN zu trainieren. Zudem ist es schwierig, bei der Suche nach den Extremwerten der Nutzenfunktion eine variable, situativ passende Schrittweitensteuerung zu realisieren (vgl. Schmidhuber 2015, Wick 2017). Diese praktischen Schwierigkeiten bei der Strukturierung wie beim Training der Netzwerke sind allesamt nur durch jeweils fallspezifische Kunstgriffe zu überwinden, die das Können ihrer Entwickler/-innen herausfordern.

Für die Gestaltung der KNN gibt es keine theoretisch fundierten Erkenntnisse. Folglich müssen für jede Aufgabe passende Netzwerkstrukturen und Nutzenfunktionen mühsam mit großem Trainingsaufwand und ohne Erfolgsgarantie ausprobiert werden. Die Performanz der Netzwerke verdankt sich daher allein der Erfahrung, dem Können und der Kreativität ihrer Entwickler/-innen, darüber hinaus auch der in den letzten Jahren gemäß dem ›Mooreschen Gesetz‹ enorm gesteigerten Leistungsfähigkeit von Computer-Hardware.

Zudem sind KNN in hohem Maße störanfällig. Schon durch geringfügige Veränderung der eingegebenen Bilddaten können sie in ihrer Funktionstüchtigkeit stark beeinträchtigt werden. So gibt es denn auch vielfältige Fehlleistungen von ansonsten erfolgreichen KNN und zahlreiche Beispiele sind belegt etwa bei der Bild-Klassifikation (vgl. z.B. Szegedy et al. 2014, Nguyen et al. 2015, Sharif et al. 2016, Sitawarin et al. 2016, Metzen et al. 2017).

(a) (b) (c) (d)

Examples of successful impersonation and dodging attacks. Fig. (a) shows SA (top) and SB (bottom) dodging against DNNB. Fig. (b)–(d) show impersonations. Impersonators carrying out the attack are shown in the top row and corresponding impersonation targets in the bottom row. Fig. (b) shows SA impersonating Milla Jovovich (by Georges Biard / CC BY-SA / cropped from https://goo.gl/GlsWIC); (c) SB impersonating SC ; and (d) SC impersonating Carson Daly (by Anthony Quintano / CC BY / cropped from https://goo.gl/VfnDct).

Abb. 2: *Ausgetrickste automatische »Gesichtserkennung«*
(Quelle: Sharif et al. 2016: o.S.)

In Anbetracht der immer häufiger eingesetzten Systeme zur automatischen ›Gesichtserkennung‹ ist das Beispiel eines mit einfachen Mitteln ausgetricksten Standardsystems zur Bildklassifikation nach neustem Stand der Technik besonders interessant und eindrücklich. Diese Systeme nutzen ebenfalls KNN, um im Falle von Gesichtsbildern anhand körperlicher Eigenheiten wie Position und Form der Nase oder Augenbrauen – mit Millionen Bildern trainiert – Personen voneinander zu unterscheiden. Werden diese Bereiche von einer Brille überdeckt, lässt das bunte Muster das KNN zur Gesichtsklassifikation Eigenheiten ausmachen, die fälschlicherweise als Gesichtsdetails ausgewertet werden. Ein männlicher

Proband wurde so als die Schauspielerin Milla Jovovich erkannt (b), mit einer Genauigkeit von 87,9 %, eine Asiatin hielt die Software mittels Brille für einen Mann aus dem arabischen Raum (c) etc. (vgl. Abb. 2).

In letzter Zeit hat das System AlphaGo von Alphabet viel Aufsehen erregt, das in Turnieren die weltbesten Go-Spieler zu schlagen vermochte. Es wird häufig und zur Überraschung vieler als der ultimative Nachweis von dem Menschen überlegener ›künstlicher Intelligenz‹ und ›maschinellem Lernen‹ präsentiert. Bei genauerem Hinsehen zeigt sich aber auch hier, dass diese Behauptungen auf dem propagandistischen Treibsand falscher Zuschreibungen gebaut sind.

Zunächst ist festzustellen, dass das Go-Spiel ein mathematisches Objekt ist, das durch seine Regeln vollständig definiert ist. Infolgedessen lässt sich jederzeit in jeder Stellung eines beliebigen Spielverlaufs eindeutig entscheiden, ob ein Spielzug erlaubt ist oder nicht. Im Prinzip ließe sich daher auch der Baum aller möglichen erlaubten Spielzüge und -verläufe darstellen (was freilich wegen sog. ›kombinatorischer Explosion‹, hier der gigantischen Zahl von geschätzt rd. 200^{150} Zweigen, physisch unmöglich ist; es handelt sich um ein NP-vollständiges Problem).

Mittels heuristischer Verfahren muss daher die Suche nach erfolgreichen Spielzügen und einem möglichst optimalen Spielverlauf auf aussichtsreiche Teilbäume beschränkt werden. Als Methode bewährt hat sich bei vergleichbaren Aufgabenstellungen die Monte Carlo Tree Search (MCTS; vgl. Browne et al. 2012), die auch hier als ein heuristischer Verfahrens-Baustein zum Tragen kommt. Dabei werden für Folgezüge einer betrachteten Stellung – dargestellt als Wurzel des Teilbaums aussichtsreicher Züge – Erfolgshäufigkeiten (Verhältnis von Anzahl gewonnener zur Anzahl der insgesamt über diesen Zweig vollzogenen Spiele) mit Hilfe von parallel durchgängig simulierten Zufallspartien ermittelt, die laufend fortgeschrieben werden (vgl. Abb. 3). Im Falle von AlphaGo wurden dazu auf 40 parallel arbeitenden Prozessoren jeweils 103 Simulationen pro Sekunde durchgeführt (was bei einer Denkzeit von einer Minute zwischen den Zügen maximal 2,4 Mio. Simulationen pro Zug ermöglicht).

In ihrer Leistung gesteigert wird die MCTS noch durch Kombination mit zwei im Spiel gegen sich selbst trainierten neuronalen Netzen, die im Spielverlauf asynchron zusätzliche Bewertungen zur Zugwahl (»policy«, in Form einer über die Zweige verteilten Erfolgswahrscheinlichkeit) und zur Stellung (»value«, als relativem Wert des zugehörigen Teilbaums) ermitteln. Dabei wächst der betrachtete Teilbaum aussichtsreicher Spielzüge durch Einführung neuer Knoten in besonders erfolgversprechenden Zweigen mit anfangs geschätzten Bewertungsgrößen. Diese werden im Zuge der parallel und asynchron durchgeführten Läufe der MCTS-Simulationen und der Wertbestimmung für Zugwahl und Stellungen durch die neuronalen Netze fortgeschrieben, sobald sie verfügbar sind (vgl. Abb. 4; weitergehende Einzelheiten der Verfahren und ihres dynamischen Zusammenspiels finden sich bei Silver et al. 2016 sowie Yuandong & Yan 2016).

↑Inhalt↑

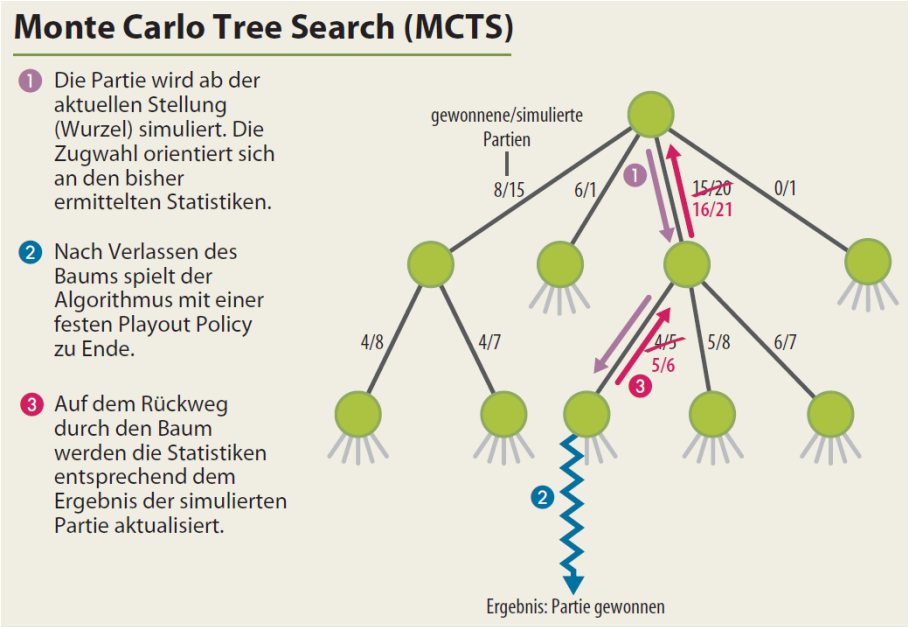


Abb. 3: *Funktionsweise der Methode Monte Carlo Tree Search (Quelle: Bögeholz 2016:151)*

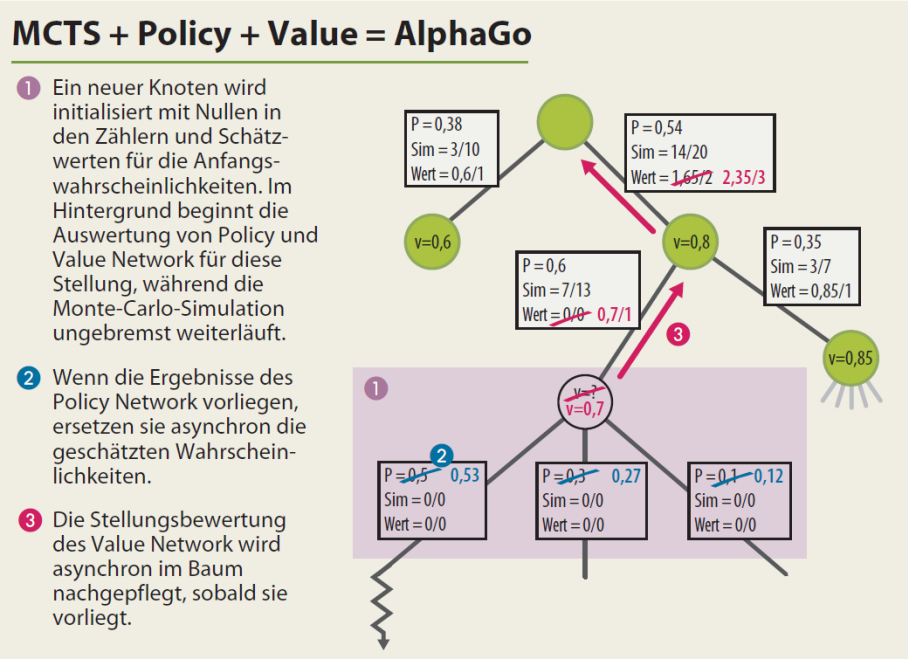


Abb. 4*: Kombination der MTCS mit KNN-basierten Bewertungen (Quelle: Bögeholz 2016: 151)*

Als mathematische, durch Regeln vollständig definierte Objekte sind Spiele entgegen landläufiger Auffassung geradezu prädestiniert für Modellierung und Formalisierung ihrer Spielverläufe. Es ist daher immer schon *a priori* sicher, dass

es Algorithmen geben muss, die menschlichen Spieler/-innen überlegen sind. Erklärungsbedürftig ist folglich nur, warum sie erst jetzt gefunden wurden; Gründe dafür sind:

- es bedarf der Entwicklung hinreichender methodischer Erfahrung und des notwendigen mathematischen Könnens, um leistungsfähige Heuristiken (z.B. MCTS) zur algorithmischen Bewältigung kombinatorischer Optimierung (oft NP-vollständiger Probleme) zu finden,
- es muss hinreichende Rechenleistung – etwa für das Training komplexer KNN – verfügbar sein.

Allgemein gilt für ›KI‹-Systeme weiterhin: Grob irreführend als ›künstlich intelligent‹ bezeichnete, *de facto* nur adaptive Computersysteme sind stets und ausschließlich das Ergebnis methodischer Kompetenz menschlicher Expert/-innen, deren Können und natürliche Intelligenz sie in Gestalt theoretischer Einsichten in zugrunde liegende Prozesse und raffiniert ausgeklügelter heuristischer Verfahren vergegenständlichen. Das gilt freilich für technische Artefakte, gleich welcher Komplexität, schon immer und manifestiert sich im Übrigen auch in der Berufsbezeichnung ›Ingenieur‹ wie im Aristotelischen *téchne* als der Kunst, etwas beruhend auf Fachwissen, Übung und Erfahrung herstellen zu können. Zudem ist, wie gezeigt, das Verhalten von KNN als adaptiven Automaten durch deren Struktur und die Vorschriften des ›Lern‹-Algorithmus zur passenden Veränderung der Gewichte vollständig determiniert (obgleich analytisch nicht mehr zu durchschauen).

3 Epistemische und ethische Anwendungsprobleme von ›KI‹-Algorithmen

Neben Herausforderungen und Schwierigkeiten der *Entwicklung* fortgeschrittener adaptiver Computersysteme bestehen aber auch auf Seiten der *Anwendung* große, wegen ihrer Besonderheit auch über den Einsatz herkömmlicher Systeme hinausweisende Probleme. Ihr Einsatz wirft gravierende ungelöste epistemische und ethische Fragen auf. Sich diesen Fragen zu stellen, ist umso dringlicher, je komplizierter und intransparenter Methoden der Modellierung von Zeichenprozessen und zugehörige algorithmische Verfahren werden und je leichter sie infolgedessen auch missbräuchlich verwendet oder manipuliert werden können (zu Details und Beispielen vgl. Autorengruppe 2018).

Infolge analytischer Intransparenz sind die Probleme adaptiver Systeme sowohl epistemischer wie ethischer Natur: Das Verhalten von ›KI‹-Algorithmen (hier v.a. KNN und Verfahren schließender Statistik) ist selbst für Entwickler/-innen aktuell weder im einzelnen durchschaubar (»inconclusive evidence«) noch im Nachhinein erklärbar (»inscrutable evidence«). Sie produzieren nur wahrscheinliche, daher stets unsichere Ergebnisse, deren Korrektheit und Validität nur schwer zu beurteilen sind. KNN sind zudem, wie oben beispielhaft gezeigt, sehr störanfällig und leicht auszutricksen. Die Ergebnisse, die sie liefern, sind in ho-

hem Maße von der Qualität der Eingabedaten abhängig, die aber meist ebenfalls unbekannt oder nur schwer einschätzbar ist (»misguided evidence«; vgl. Mittelstadt et al. 2016: 4f).

Nutzer/-innen können dem Verhalten und seinen Ergebnissen daher nur blind vertrauen – trotz der nicht aufhebbaren Unsicherheit. Das stellt sie in der »Koaktion« (vgl. Hubig in diesem Band) mit solchen Systemen vor beträchtliche Belastungen: Wie sollen sie sich solche adaptiven Systeme überhaupt aneignen, wie mit ihnen zweckmäßig und zielgerichtet koagieren, wenn diese sich in vergleichbaren Situationen jeweils anders und unerwartet verhalten? Das wäre ein eklatanter Verstoß gegen eine der Grundregeln der Mensch-Maschine-Interaktion, gegen die Forderung nach erwartungskonformem Verhalten (vgl. EN ISO 9241-11 Anforderungen an die Gebrauchstauglichkeit). Zugleich würden auf Seiten der Nutzer/-innen stets aufs Neue überzogene Erwartungen an die vermeintliche Leistungsfähigkeit der Systeme geschürt, mithin gar ihre Wahrnehmung der Wirklichkeit verändert (»transformative effects«, Mittelstadt et al. 2016: 5). Konfrontiert mit diesen Widersprüchen, unter dem Erwartungsdruck erfolgreicher Bewältigung ihrer Aufgaben einerseits und angesichts des Verlusts der Kontrolle über Arbeitsmittel mit undurchschaubarem Verhalten andererseits, würden sie unter dauerhaften psychischen Belastungen zu leiden haben (so bereits Norman 1994 im Hinblick auf ›Agenten‹ als frühen adaptiven Systemen).

Darüber hinaus stellen sich mit diesen Systemen auch Fragen nach der ethischen Verantwortbarkeit: Dürfen derart undurchschaubare und störanfällige Artefakte überhaupt in die Welt gesetzt werden, da ihr künftiges Verhalten nicht sicher vorhersehbar ist? Wer ist gegebenenfalls für eingetretene Schäden verantwortlich – Entwickler/-innen?, Betreiber/-innen? oder Nutzer/-innen? – und wie werden daraus entstehende Haftungsansprüche geregelt? Bisher getroffene oder sich abzeichnende Regelungen sind unbefriedigend und unzureichend.

Auch hier gelten seit langem diagnostizierte »Ironien der Automatisierung« (Bainbridge 1983) zugespitzt weiter: Von besonderer Bedeutung ist dabei die Ironie, dass mangels hinreichender Übung und Erfahrung bei automatischem Normalbetrieb ausgerechnet die im Stör- oder Versagensfall wiederum benötigte menschliche Handlungskompetenz schwindet. Es fehlen Ansätze, wie dieser Art ›erlernter Inkompetenz‹ entgegengewirkt werden kann. In Schadensfällen wird die

Ursache meist dichotomisch in ›menschlichem Versagen‹ oder einer technischen Störung gesucht und dabei die wahre, in fehlgeleiteter soziotechnischer Systemgestaltung liegende Ursache ignoriert. Die derart systemisch bedingte mangelnde Kompetenz, in Verbindung mit der durch Untätigkeit oder Ablenkung geschwächten Vigilanz, führt dann bei plötzlich notwendigen Eingriffen zu beträchtlichen Problemen der Bewältigung:

- unzureichende Erfahrungen und verlernte Fähigkeiten können zu fehlerhaften Diagnosen und falschen oder riskanten Aktionen führen,

- es entstehen, wie die Empirie ergibt, lange Verzögerungszeiten von 7 bis 10 Sekunden bis zum Zurechtfinden in der unerwarteten und ungewohnten Situation und zur Rückgewinnung von Handlungsfähigkeit (ein 150 km/h schnelles Auto fährt in dieser Zeit rd. 400 m weit).

Beispiele und empirische Erkenntnisse zu diesen Herausforderungen gibt es aus der Forschung über Leitwartentätigkeiten, Flugführung oder automatisiertes Fahren zuhauf, gleichwohl wurden bislang wenig weiterführende Konsequenzen gezogen. Daher ist infolge dieser noch weitgehend ungelösten Schwierigkeiten im praktischen Einsatz und Gebrauch adaptiver Systeme in naher Zukunft mit beträchtlichen Verzögerungen der Entwicklung zu rechnen (vgl. Bainbridge 1983, Baxter et al. 2012, Casner/Hutchins/Norman 2016, Casner et al. 2016, DIVSI 2016, Weyer 2007).

4 Historischer Exkurs: Meilensteine der Entwicklung von Computertechnik und Computing Science

Die hier angesprochenen Schwierigkeiten und Herausforderungen von Entwicklung und Anwendung fortgeschrittener adaptiver Computersysteme werden im landläufigen Verständnis von Computertechnik weitgehend ignoriert. Dadurch werden Fallstricke der Realisierung verkannt, Potenziale maßlos übertrieben und gesellschaftliche Folgen falsch eingeschätzt. Dies ist aber beileibe keine neue Erscheinung, sondern begleitet die Computertechnik seit dem Beginn ihrer materiellen Manifestation, die gemeinhin mit rüstungstechnischen Entwicklungen im und kurz nach dem 2. Weltkrieg angesetzt werden. Seither beherrschen prinzipiell irreführende, aber euphorisierende Metaphern wie ›Elektronengehirn‹, ›künstliche Intelligenz‹, ›maschinelles Lernen‹ oder ›autonome Systeme‹ bis hin zu den eingangs zitierten Äußerungen den gesellschaftlichen, oft aber auch den wissenschaftlichen Diskurs. Nur gelegentlich, in Zeiten großer Ernüchterung angesichts wirklicher Probleme der Modellierung und Algorithmisierung von Zeichenprozessen sozialer Praxis, ist weit zutreffender etwa von ›elektronischer Datenverarbeitung‹ die Rede.

In der ansonsten irreführenden, durchweg anthropomorphisierenden Metaphorik kommt aber ein grundsätzlich fehlgeleitetes Verständnis von Computertechnik zum Ausdruck. Computersysteme führen, wie die Theorien der Computing Science lehren, mittels Daten als auf Syntax reduzierten Zeichen berechenbare Funktionen aus und sonst nichts. Ihr Verhalten ist durch die implementierten Algorithmen und eingegebene Daten vollständig determiniert (nach dem Modell der Turingmaschine). Buchstäblich ›wissen‹ sie nicht, was sie tun. Intelligent sind daher nicht die Computersysteme, sondern ausschließlich ihre Entwickler/-innen, die zuvor in mühevoller Arbeit mit ihrer Kreativität und methodischen Kompetenz Zeichenprozesse sozialer Praxis nach gewünschten Anforderungen modelliert, formalisiert und in berechenbare Funktionen überführt haben, oder ggf. auch der-

en Nutzer/-innen, die durch Aneignung ihrer Funktionen damit etwas Sinnvolles anzustellen vermögen (nähere Einzelheiten hierzu finden sich, gestützt auf den triadischen Zeichenbegriff von Peirce (1983), bei Brödner 2008, 2018 sowie Nake 2001). Diese professionelle Sicht wurzelt in der geschichtlichen Entwicklung von Computertechnik und Computing Science (die Bezeichnung ›Informatik‹ ist ebenfalls eine irreführende Fehlbenennung: es geht um Verarbeitung von Daten, nicht von ›Information‹; vgl. Brödner 2014), wie sie in ihren logischen und konzeptionellen historischen Meilensteinen zum Ausdruck kommt (vgl. Übersicht).

Tatsächlich beginnt die Entwicklung nicht erst im 2. Weltkrieg, sondern bereits in der Frühphase der industriellen Revolution mit der hochgradig arbeitsteiligen Organisation von Kopfarbeit, mittels derer anspruchsvolle kognitive Aufgaben (etwa die Neuberechnung umfangreicher mathematischer Tafeln im Dezimalsystem, z.B. Logarithmen, nautische Almanache, Artillerie-Schusstafeln) mittels Formularen in eine geplante Abfolge einfachster Rechenoperationen (Addition und Subtraktion reeller Zahlen mit Hilfe mechanischer Rechenmaschinen) aufgelöst und zur Ausführung vorgeschrieben werden. Dies geschieht in enger ideeller Verzahnung mit der arbeitsteiligen Organisation von Handarbeit in kleinste wiederkehrende Verrichtungen und ihrer anschließenden Mechanisierung mittels Arbeitsmaschinen (z.B. Textilmaschinen, Werkzeugmaschinen). Im Laufe der Entwicklung werden Funktionen des *Antriebs* (›Kraftmaschinen‹), der *Werkzeugführung* und der *Steuerung* zunehmend voneinander getrennt; dadurch werden für Antriebe fossile Energie (Dampfmaschinen mit Transmission, später Einzelantrieb mit Elektromotoren) nutzbar und Steuerungen realisiert, die statt Kräften Signale (Daten) über maschinelle Bewegungen verarbeiten, um deren gewünschten Ablauf zu gewährleisten. In dieser Perspektive lassen sich Computer auch als *universelles Steuerungspotenzial* begreifen, das per Programm in eine spezifische Steuerung verwandelt wird (wie es heute bei digitaler Prozesssteuerung üblich ist).

Die arbeitsteilige Organisation spezialisierter Verrichtungen erfordert zunehmend aufwendigere Planung und sachliche wie zeitliche Koordination der Einzelarbeiten durch ›Manager/-innen‹. Wachsender Aufwand für Koordination und Sicherung der Herrschaft über immer kompliziertere Prozesse der (Massen-)Produktion erfordern zudem Maßnahmen vereinfachender Standardisierung sowie wissensbasierter Planung, Anweisung und Kontrolle (Taylors Prinzipien des »Scientific Management«). Diese vertikale Arbeitsteilung der Trennung von Planung und Ausführung beruht auf expliziten Beschreibungen von Produkten und Prozessen, führt mithin zu einer ›Verdoppelung‹ der Produktion in Zeichen (in Form von Zeichnungen, Stücklisten, Arbeitsplänen etc.). Das resultiert insgesamt in fortschreitender Verwissenschaftlichung von Produktion: Mit der Analyse, Planung und Kontrolle von Produktionsprozessen wird laufend erweitertes explizites Wissen über sie gewonnen und mit anderen wissenschaftlichen Erkenntnissen kombiniert. Dieses Wissen wächst wie ein Baum durch Verzweigung und wird durch Zeichen repräsentiert. So entstehen mit der ›Verdoppelung‹ der Arbeitswelt

in Zeichen auch zunehmend durch Zeichen repräsentierte Arbeitsgegenstände und ebensolche Methoden ihrer Verarbeitung. Die Anwendung dieses expliziten propositionalen Wissens zur Lösung wirklicher praktischer Probleme erfordert allerdings in wachsendem Maße wiederum Können, Wissensteilung und Kooperation von spezialisierten Wissensarbeiter/-innen.

***Übersicht 1**: Meilensteine der Entwicklung von Computertechnik und Computing Science*

1792-1801:	*Gaspard de Prony* entwickelt ein formularbasiertes Verfahren zur extrem arbeitsteiligen Neuberechnung mathematischer Tafeln im Dezimalsystem; das Formular-Schema der Abfolge einfacher Rechenoperationen bildet die Urform eines Algorithmus (noch im 2. Weltkrieg wurden V2-Flugbahnen so berechnet).
1805:	Jacquard-Webstuhl, erste digital mittels ›Lochbrettern‹ gesteuerte Arbeitsmaschine.
1812:	Charles Babbage konzipiert die Difference Engine zur Berechnung der Funktionswerte von Polynomen: $f(x) = a_n x^n + \ldots + a_1 x + a_0$ (teils öffentlich gefördert, 1822 realisiert).
Um 1830:	Charles Babbage entwirft und programmiert die allgemeiner verwendbare Analytical Engine. Sie nimmt die von-Neumann-Architektur programmierbarer Universalrechner (Rechenwerk, Speicher, Steuerung, Datenein- & -ausgabe) vorweg, scheitert aber an der mechanischen Realisierung.
1847:	George Boole publiziert einen Logikkalkül (um 1888 von Peano als Boolesche Algebra axiomatisiert); er bildet das logisch-funktionale Fundament für binäre Schaltsysteme (als Kern heutiger Computer-Hardware).
1860-1880:	Charles S. Peirce entwickelt erstmals einen Prädikatenkalkül 1. Stufe, arbeitet an ›logischen Maschinen‹ und entwickelt die bislang elaborierteste Theory of Signs (triadische Zeichentheorie, ohne die computerisierte Kopfarbeit gar nicht zu verstehen wäre).
1931:	Kurt Gödel beweist die Unvollständigkeit formaler Systeme wie das der *principia mathematica* von A.N. Whitehead & B. Russell.
1936:	Alan Turing publiziert die Idee der Turingmaschine und definiert damit formal die Begriffe Algorithmus und berechenbare Funktion (äquivalent: Lambda-Kalkül von Church & Kleene 1936).
1945:	Konrad Zuse entwickelt den Plankalkül als erste algorithmische Programmiersprache (in Anlehnung an den Lambda-Kalkül).

Erst auf der Grundlage dieser Entwicklungsgeschichte der Computertechnik lässt sich ermessen, wie und warum Computer ihren Siegeszug durch die Arbeitswelt antreten konnten, sobald erst einmal mit elektro-mechanisch, später elektronisch realisierten binären Schaltsystemen die passende Hardware zur Verarbeitung binär codierter Daten als auf Syntax reduzierten Zeichen gefunden war.

Im Zuge der ganzen Entwicklung bilden bis heute die logisch-konzeptionellen Ideen der Software mit den auf Erfahrung, Kreativität und Können ihrer Entwickler/-innen, auf deren »lebendiges Arbeitsvermögen« (Pfeiffer 2004) angewiesenen Modellierungsmethoden und algorithmischen Verfahren den führenden Faktor und die Leistung der Hardware das limitierende Nadelöhr.

5 Prinzipielle Grenzen und Missverständnisse der Computertechnik

In diesem Zusammenhang ist zunächst an prinzipielle Grenzen der Formalisierung von Zeichenprozessen zu erinnern. Selbst die äußerst formalisierte Mathematik widersetzt sich ihrer vollständigen Algorithmisierung. Ausgerechnet im Zusammenhang mit den zu Beginn des 20. Jahrhunderts bestehenden großen Hoffnungen auf eine vollständige Formalisierung der bekannten Mathematik hat sich herausgestellt, dass es erwiesenermaßen unmöglich ist,

- einen Algorithmus anzugeben, der alle Sätze eines formalen Systems abzuleiten und deren Widerspruchsfreiheit zu zeigen imstande ist (vgl. Gödel 1931);
- einen Algorithmus anzugeben, der von jeder Formel eines formalen Systems entscheiden kann, ob diese Formel ein wahrer Satz des Systems ist (vgl. Turing 1936).

Bezeichnenderweise beruht der Beweis von Gödel im Kern darauf, dass er als erfahrener und kompetenter Mathematiker eine Formel im System so zu konstruieren vermag, 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 mit deren Hilfe formalisieren können, durch Nachdenken über die Formalisierung mittels abduktiven Schließens über sie hinaus zu gelangen vermögen (vgl. Brödner 1997, 2018).

Als rein formales Verfahren gibt die Abfolge von Operationen eines Algorithmus zwar Auskunft auf die Frage, was genau operativ abläuft; sie beantwortet aber nicht die Frage nach deren Sinn oder Bedeutung, warum sie so abläuft – eben deshalb gehört zur Software auch die Dokumentation mit derartigen Meta-Aussagen. Mit rein formalen Mitteln gelingt es eben nicht, auf der Metaebene Aussagen über Terme auf der operativen Ebene zu machen. Operationsfolgen sagen nichts über sich selbst aus, etwa ob sie korrekt oder gebrauchstauglich sind. So ist etwa auch die Frage, ob ein Algorithmus terminiert, formal nicht entscheidbar.

Das aus dem geschichtlichen Werdegang von Computertechnik und Computing Science gewonnene Verständnis der Funktionsweise von Computern als semiotischen Maschinen erlaubt zudem, die fundamentalen Unterschiede zu Menschen als lebendigen Organismen aufzuzeigen (vgl. Übersicht 2). Die ständige Rede von ›künstlich intelligenten‹ oder gar ›autonomen‹ Computersystemen entpuppt sich dabei als folgenreicher Etikettenschwindel. Wieder einmal

bedarf es der Philosophie als »Kampf gegen die Verhexung unseres Verstandes durch die Mittel unserer Sprache« (Wittgenstein 1984: PU 109).

Die irreführende Metaphorik über Computer, als ob diese ›wie Menschen‹ intentional eingestellt und handlungsfähig wären – ›autonom‹, ›selbstorganisiert‹, ›intelligent‹, ›smart‹, ›selbstlernend‹, ›selbstheilend‹ etc. –, ignoriert in reduktionistischer Weise nach Denkmustern des Funktionalismus die fundamentalen Unterschiede. Dadurch wird kompetentes Handeln von Menschen auf algorithmisch gesteuertes Verhalten von Maschinen reduziert; zugleich entstehen eben dadurch Illusionen über deren Zustandekommen und tatsächliche Leistungsfähigkeit. Das führt im Ergebnis zu einer verbreiteten Selbsttäuschung, wie sie in Diskursen um ›künstliche Intelligenz‹ zum Ausdruck kommt (Brödner 2018). In den Wahnvorstellungen vom vermeintlichen Eigenleben der Maschinen äußert sich deren Fetischcharakter, die »Macht der Machwerke über die Machenden« (Haug 2005: 162).

***Übersicht 2**: Ontologische Differenz zwischen Mensch und Computer*

Mensch (lebendiger Organismus)	**Computer** (semiotische Maschine)
Sich durch Autopoiese in Stoffwechsel und Kommunikation selber machend.	Wissensbasiert für bestimmte Zwecke gemacht (konstruiert).
Autonom (selbstbestimmte Regeln).	Automatisch (auto-operational, selbsttätig).
Handelt intentional (kontingent).	Verhält sich kausal determiniert;
Ist sprachbegabt, reflektiv lernfähig.	ggf. algorithm. gesteuert Umwelt-adaptiv.
Lebendiges Arbeitsvermögen: Können (implizites Wissen, Erfahrung, situierte Urteilskraft & Handlungskompetenz),verausgabt & reproduziert sich im Gebrauch.	Algorithmisch determiniertes Verhalten: Setzt Formalisierung von Zeichenprozessen voraus, muss für die Praxis angeeignet & organisatorisch eingebettet werden.

6 Fazit: Grenzen und Widersprüche adaptiver Systeme

Aus diesen Ausführungen können mit Blick auf Grenzen und Widersprüche der Entwicklung und Anwendung fortgeschrittener adaptiver Computersysteme unmittelbar verschiedene Schlussfolgerungen gezogen werden:

Erstens bestimmen sog. ›autonome Systeme‹ die Regeln ihres Verhaltens nicht selbst, folglich sind sie tatsächlich nicht autonom, sondern als adaptive Automaten konstruiert.

Zweitens liegt es in der Regel-Natur von Spielen, ihrer Natur als mathematischem Objekt, dass Algorithmen – bei hinreichender Leistung der Hardware – menschlichen Spieler/-innen überlegen sind. Aus dieser Besonderheit kann aber

nicht allgemein auf die Überlegenheit maschineller Verfahren über menschliche kognitive Kompetenz geschlossen werden.

Drittens gelingt es menschlicher Kreativität immer wieder, für spezielle, auch schwierige Aufgaben der Zeichenverarbeitung algorithmische Lösungsverfahren zu finden und als adaptive Automaten zu realisieren, die aber nur sehr begrenzt auf andere Aufgaben übertragbar sind. Meist ist die aufwendige Entwicklung jeweils eigener Methoden erforderlich, oft mit mehr Aufwand als Nutzen. Sog. ›KI‹-Verfahren sind daher keine »General Purpose Technology«, wie das Brynjolfsson et al. (2017: 19f) leichtfertig behaupten.

Viertens beruhen ironischerweise die Strukturen und Algorithmen erfolgreich eingesetzter adaptiver Systeme – mangels theoretischer Einsichten in Ursache-Wirkungs-Ketten – ausschließlich auf der Erfahrung und Kreativität, mithin dem Können bzw. dem Arbeitsvermögen ihrer Entwickler/-innen.

Fünftens ist das Verhalten adaptiver Systeme aus gleichem Grund nur schwer oder gar nicht zu durchschauen oder zu erklären, zudem äußerst störanfällig. In Form von KNN oder Verfahren schließender Statistik liefern sie stets nur wahrscheinliche, daher prinzipiell unsichere Ergebnisse. Das macht instrumentelles Handeln mit ihnen schwierig bis unmöglich, jedenfalls psychisch hoch belastend – gefordert sind daher Transparenz und Kontrolle ihrer algorithmischen Funktionen durch unabhängige Institutionen (wie bei anderer hoch riskanter Technik auch).

Statt auf illusionäre ›KI‹-Hoffnungen zu setzen, erscheint es angesichts dieser Einsichten in die zweifelhafte Tragfähigkeit von Konzepten ›künstlicher Intelligenz‹ weit aussichtsreicher, höhere Flexibilität, Produktivität und Innovationsfähigkeit stattdessen durch soziotechnische Gestaltung »guter Arbeit« zu erreichen. Dazu notwendiges Wissen ist aus über drei Jahrzehnten Forschung zur Gestaltung von Arbeit und Technik verfügbar. Dies erscheint umso notwendiger, je mehr die künftige gesellschaftliche Entwicklung auf die breite Entfaltung menschlichen Arbeitsvermögens für den produktiven Umgang mit komplexen Beständen explizten Wissens und technischen Systemen angewiesen ist. Gestützt auf dieses Wissen ließe sich das immer bedeutender werdende lebendige Arbeitsvermögen durch Organisation reflexiven und kreativen Zusammenwirkens kompetenter Expert/-innen mit gebrauchstauglich gestalteten Computersystemen weit wirksamer als durch den Einsatz von ›KI‹-Systemen zur Entfaltung bringen.

7 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

Babbage, Charles (1832): On the Economy of Machinery and Manufactures, London: Knight, Reprint New York: Kelley 1971 (deutsch: Die Ökonomie der Maschine, Nachdruck der Originalübersetzung von 1833, hg. von P. Brödner, Berlin: Kulturverlag Kadmos 1999)

Bainbridge, Lisanne (1983): »Ironies of Automation«, in: Automatica 19 (6), S. 775-779.

Bastian, Matthias (2018): »Google-Chef: Künstliche Intelligenz ›wichtiger als Feuer und Elektrizität‹«, in: Vrodo vom 20.01.2018, https://vrodo.de/google-chef-kuenstliche-intelligenz-wichtiger-als-feuer-und-elektrizitaet/

Baxter, Gordon/Rooksby, John/Wang, Yuanzhi/Khajeh-Hosseini, Ali (2012): »The Ironies of Automation ... still going strong at 30?«, in: Phil Turner/Susan Turner (Hg.): European Conference on Cognitive Ergonomics, ECCE '12, Edinburgh, United Kingdom, August 28 - 31, 2012, S. 65-71.

Bögeholz, Harald (2016): »Mysteriöse Tiefe. Wie Google-KI den Menschen im Go schlagen will«, in: c't 6/2016, S. 148-151.

Brödner, Peter (2018): »Coping with Descartes' Error in Information Systems«, AI & Society Journal of Knowledge, Culture and Communication 2018 (online first).

Brödner, Peter (2014): »Durch „Information" desinformiert? Zur Kritik des Paradigmas der Informationsverarbeitung«, Arbeits- und Industriesoziologische Studien 7 (1), S. 42-59

Brödner, Peter (2008): »Das Elend computerunterstützter Organisationen«, in: Dorina Gumm/Monique Janneck/Roman Langer/Edouard J. Simon (Hg.): Mensch –Technik – Ärger? Zur Beherrschbarkeit soziotechnischer Dynamik aus transdisziplinärer Sicht, Münster: Lit-Verlag, S. 39-60.

Brödner, Peter (1997): Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen, Berlin: edition sigma.

Browne, Cameron/Powley, Edward/Whitehouse, Daniel/Lucas, Simon/Cowling Peter I./Rohlfshagen, Philipp/ Taverner, Stephen/Perez, Diego/Samothrakis, Spyridon/Colton, Simon (2012): »A Survey of Monte Carlo Tree Search Methods«, in: IEEE Transactions on Computational Intelligence and AI in Games 4 (1), S. 1-49.

Brynjolfsson, Erik/Rock, Daniel/Syverson, Chad (2017): Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, NBER Working Paper No. 24001.

Casner, Stephen M./Geven, Richard W./Recker, Matthias P./Schooler, Jonathan W. (2014): »The Retention of Manual Flying Skills in the Automated Cockpit«, in: Human Factors 56 (8), S. 1506-1516.

Casner, Stephen M./Hutchins, Edwin L./Norman, Don (2016): »The Challenges of Partially Automated Driving«, CACM 59 (5), S. 70-77.

Deutsches Institut für Vertrauen und Sicherheit im Internet (DIVSI) (2016): Digitalisierte urbane Mobilität. Datengelenkter Verkehr zwischen Erwartung und Realität, Hamburg: Deutsches Institut für Vertrauen und Sicherheit im Internet.

Dworschak, Manfred (2018): »Was künstliche Intelligenz schon leisten kann - und was nicht«, in: Der Spiegel 2/2018

Gödel, Kurt (1931): Über formal unentscheidbare Sätze der principia mathematica und verwandter Systeme I, Monatshefte für Mathematik und Physik 38 (1), S. 173-198.

Haug, Wolfgang F. (2005): Vorlesungen zur Einführung ins »Kapital«, Hamburg: Argument.

Hilberg, Wolfgang (1995): »Karl Steinbuch – ein zu Unrecht vergessener Pionier der künstlichen neuronalen Systeme«, in: Frequenz 49 (1-2), S. 28-36.

LeCun, Yann/Bengio, Yoshua/Hinton, Geoffrey (2015): »Deep Learning«, in: Nature 521, S. 436-444.

LeCun, Yann/Bottou, Leon/Bengio, Yoshua/Haffner, Patrick (1998): »Gradient-Based Learning Applied to Document Recognition«, Proceedings of the IEEE 11, S. 2278-2324.

McCulloch, Warren S/Pitts, Walter H. (1943): A Logical Calculus of the Ideas Immanent in Nervous Activity, Bulletin of Mathematical Biophysics 5, S. 115-133.

Metzen, Jan H./Kumar, Mummadi C./Brox, Thomas/Fischer, Volker. (2017): Universal Adversarial Perturbations Against Semantic Image Segmentation, arXiv:1704.05712v1, https://arxiv.org/pdf/1704.05712v1.pdf

Minsky, Marvin/Papert, Seymour (1969): Perceptrons, Cambridge, MA: MIT Press

Mittelstadt, Brent D./Allo, Patrick/Taddeo, Mariarosaria/Wachter, Sandra/Floridi, Luciano (2016): »The Ethics of Algorithms: Mapping the Debate«, in: Big Data & Society 3 (2), S. 1-21.

Nake, Frieder (1992): »Informatik und Maschinisierung von Kopfarbeit«, in: Wolfgang Coy/Frieder Nake/Jörg-Martin Pflüger/Arno Rolf/Jürgen Seetzen/Dirk SiefkesReiner Stransfeld. (Hg.): Sichtweisen der Informatik, Braunschweig/ Wiesbaden: Vieweg, S. 181-201.

Nguyen, Anh/Yosinski, Jason/Clune, Jeff (2015): »Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images«, in: Computer Vision and Pattern Recognition (15), IEEE, https://arxiv.org/pdf/1412.1897.pdf.

Norman, Donald A. (1994): »How Might People Interact with Agents«, in: CACM 37 (7), S. 68-71.

Ortega y Gasset, José (1949): Betrachtungen über die Technik, Stuttgart: DVA

Peirce, Charles S. (1983): Phänomen und Logik der Zeichen, Frankfurt/M: Suhrkamp (engl. Original: A Syllabus of Certain Topics of Logic, Boston Alfred Mudge & Son, 1903).

Recke, Martin (2016): »Invasion der Roboter: Künstliche Intelligenz ist bald so normal wie Strom«, in: t3n vom 13.03.2016, https://t3n.de/news/sxsw-traurige-roboter-688398/

Ropohl, Günter (1991): Technologische Aufklärung. Beiträge zur Technikphilosophie, Frankfurt/M: Suhrkamp.

Schäfer, Ulrich (2016): »Leere Büros, leere Fabriken«, in: Süddeutsche Zeitung vom 10.11.2016, http://www.sueddeutsche.de/wirtschaft/kuenstliche-intelligenz-leere-bueros-leere-fabriken-1.3243399

Schmidhuber, Jürgen (2015): »Deep Learning in Neural Networks. An Overview«, in: Neural Networks 61, S. 85-117.

Sharif, Mahmood/Bhagavatula, Sruti/Bauer, Lujo/Reiter, Michael K. (2016): Accessorize to a Crime. Real and Stealthy Attacks on State-of-the-Art Face Recognition,in: CCS '16 Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, https://users.ece.cmu.edu/~mahmoods/publications/ccs16-adv-ml.pdf (Abruf: 18.02.2018).

Silver, David/Huang, Anja/Maddison, Chris J./Guez, Arthur/Sifre, Laurent/van den Driessche, George/Schrittwieser, Julian/Antonoglou, Ioannis/Panneershelvam, Veda/Lanctot, Marc/Dieleman, Sander/Grewe, Dominik/Nham, John/ Kalchbrenner, Nal/Sutskever, Ilya/Leach, Madeleine/Kavukcuoglu, Koray/Graepel, Thore/Hassabis, Demis (2016): »Mastering the Game of Go with Deep Neural Networks and Tree Search«, in: Nature 529, S. 484-489.

Simon, Hubert A./Newell, Allen (1972): Human Problem Solving, Englewood Cliffs (NJ): Prentice Hall.

Sitawarin, Chawin/Bhagoji, Arjun N./Mosenia, Arsalan/Chiang, Mung/Mittal, Prateek (2018): DARTS: Deceiving Autonomous Cars with Toxic Signs, arXiv:1802.06430v1, https://arxiv.org/pdf/1802.06430.pdf

Sulleyman, Aatif (2017): »Stephen Hawking warns Artificial Intelligence ›may replace humans altogether‹ «, in: Independent vom 02.11.2017, https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawkingartificial- intelligence-fears-ai-will-replace-humans-virus-life-a8034341.html

Szegedy, Christian/Zaremba, Wojciech/Sutskever, Ilya/Bruna, Joan/Erhan, Dumitru/Goodfellow, Ian/Fergus, Rob (2014): Intriguing Properties of Neural Networks, arXiv:1312.6199v4, https://arxiv.org/pdf/1312.6199.pdf

Trinkwalder, Andrea (2016): »Netzgespinste. Die Mathematik neuronaler Netze: Einfache Mechanismen, komplexe Konstruktion«, in: c't 6/2016, S. 130-135.

Turing, Alan (1936): »On Computable Numbers. With an Application to the Entscheidungsproblem«, in: Martin Davis(Hg.): The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems, and Computable Functions, New York: Raven 1965, S. 116-151.

Verein Deutscher Ingenieure (Hg.) (1991): Technikbewertung – Begriffe und Grundlagen. Erläuterungen und Hinweise zur VDI-Richtlinie 3780, VDI Report 15, Düsseldorf: VDI.

Weyer, Johannes (2007): »Autonomie und Kontrolle. Arbeit in hybriden Systemen am Beispiel der Luftfahrt«, in: Technikfolgenabschätzung – Theorie und Praxis 16 (2), S. 35-42.

Wick, Christoph (2017): Deep Learning, Informatik Spektrum 40 (1), S. 103-107.

Wittgenstein, Ludwig (1984): Philosophische Untersuchungen. Werkausgabe Bd. 1, Frankfurt/M: Suhrkamp: S. 225-580.

Yuandong Tian/Yan, Zhu (2016): Better Computer Go Player with Neural Network and Long-term Prediction, arXiv: 1511.06410v3, https://arxiv.org/pdf/1511.06410.pdf

↑Inhalt↑

»Machines that think« – die »KI«-Illusion und ihre Wurzeln

»Wer denkt, dass eine Maschine denkt, denkt wie eine Maschine.«
(F. Nake)

Im gegenwärtigen, gelegentlich Züge von Hysterie aufweisenden gesellschaftlichen Diskurs um »Digitalisierung« und »Künstliche Intelligenz (KI)« werden die Gegenstände in einer Weise bezeichnet, die diese in hohem Maße mystifizieren und umnebeln, ihre wahre Natur mithin weitgehend im Dunkeln lassen. Das beginnt schon mit dem gänzlich undifferenzierten Gebrauch der Bezeichnung »Digitalisierung« und setzt sich fort in den zumeist märchenhaften Erzählungen über vermeintliche »KI«. Am Ende leidet sogar die Fachdisziplin »Informatik« selbst darunter, dass ihr eigentlicher Gegenstand und die Kernfragen seiner Analyse, Gestaltung und Bewertung unzureichend geklärt erscheinen. Dem wird im folgenden in mehreren Schritten nachgegangen, um zu einem vertieften Verständnis der hinter den »KI«-Ansprüchen liegenden Illusionen zu gelangen.

1 Falsche Bezeichnungen führen das Denken in die Irre

Bereits vom Beginn ihrer Realisierung an beruhen häufige Darstellungen der Berechnungsleistungen von Computern auf meist falschen, daher irreführenden Metaphern. Eine kleine Auswahl von Zeitungsmeldungen aus der Frühzeit der technischen Realisierung von Computern mag das illustrieren:

- »Elektronen›gehirn‹ berechnet 100-Jahres-Problem in 2 Stunden.« (New York Harald Tribune 15.02.1946),
- »30-Tonnen-Elektronengehirn an der Philadelphia Universität denkt schneller als Einstein.« (Philadelphia Evening Bulletin 15.02.1946),
- »Giant Brains or Machines that Think« (Buchtitel von E.C. Berkeley 1949),
- ›Time‹- Titelbild mit dem Computer Mark III: »Can Man Build a Superman?« (Time Magazine 23.01.1950).

Damit werden Programme, die mehr oder weniger komplizierte Berechnungsverfahren formal beschreiben und in ihrem Ablauf steuern, leichthin mit intelligentem Handeln von Menschen gleichgesetzt. Die gravierenden Folgen dieser unpassenden Metaphorik und des darin zum Ausdruck gebrachten ›Zeitgeists‹ zeigen sich beispielsweise

- im Turing-Test (1950), der – geboren aus dem Geist des Behaviorismus – eben darauf beruht: programmgesteuertem Verhalten von Computern wird

intelligentes menschliches Handeln zugeschrieben, falls Versuchspersonen in einem schriftlich geführten Dialog nicht mehr zu unterscheiden vermögen, ob die jeweiligen Antworten von einem Menschen oder einem Computer stammen;

- im von McCarthy (1955) ausgerufenen Forschungsprojekt zur »Artificial Intelligence (AI)«, die er als das Problem definiert: »... making a machine behave in ways that would be called intelligent if a human were so behaving.«

Diese Wortschöpfung war, nebenbei bemerkt, ein genialer Propaganda-Coup: Wurde die Entwicklung vergleichbar komplexer mathematischer Methoden und Berechnungsverfahren zuvor unter der apokryphen Fachbezeichnung »Operations Research (OR)« behandelt (die nur Experten interessierten), fanden sie als »AI« (für »KI«) nun plötzlich große öffentliche Aufmerksamkeit (obgleich das Projekt kaum greifbare Ergebnisse erbrachte, außer dem Anstoß zur späteren funktionalen Programmiersprache LISP (McCarthy 1960)). So entsteht etwa, beruhend auf dieser suggestiven, aber durch nichts weiter gerechtfertigten Gleichsetzung, das sog. »computational model of the mind«, das bis heute die Kognitionswissenschaften beherrscht. Es tritt in zwei Varianten auf:

- als ›schwache KI‹: Computer ahmen intelligentes Verhalten von Menschen lediglich nach, oder auch
- als ›starke KI‹: »Cognition is Computation« (Pylyshyn 1984).

Diese durchweg irrtümlichen Zuschreibungen bedürfen dringend der Korrektur, da sie zum einen die Genese und Funktionsweise von Computern ebenso missverstehen wie sie zum anderen die Leistungen des menschlichen Verstandes verkennen.

Keineswegs besser steht es auch um die ubiquitär gebrauchte, aber gänzlich irreführende Bezeichnung »Digitalisierung« für Vorgänge, bei denen es tatsächlich um Entwicklung und Einsatz von Computersystemen (früher: »elektronische Datenverarbeitung«) als ›Werkzeugen‹ und deren weltweite Vernetzung zu einem Medium der Kommunikation und Kooperation geht. Als Zeichen gebrauchendes »semiotisches Tier« (Hausdorff 1897) vermag der homo sapiens auf sehr verschiedene Weise mittels Zeichen Abbildungen oder Beschreibungen seiner Welt zu erzeugen: Analog heißen sie, wenn sie abbildender Natur sind und zu Anschauungen führen; digital werden sie genannt, wenn sie sprachbasiert sind und mittels Worten aus diskreten Zeichen (Buchstaben und Ziffern) der Bezeichnung dienen und Begriffe darstellen (abgeleitet von *digitus*, dem Finger als frühem Zahlzeichen; entsprechend ist digital auch das Kennzeichen eines zeit- und wertdiskreten Signals).

Schrift- und Zahlzeichen sind aber seit rd. 7.000 Jahren in Gebrauch (mit zwischenzeitlich beachtlichen technisch-medialen Errungenschaften); wie kann dann mit »Digitalisierung« irgend etwas Spezifisches über derzeit zweifellos gewichtige technische Entwicklungen bezeichnet werden? Tatsächlich gemeint ist wohl – so lässt der Gebrauchskontext vermuten – »Computerisierung«, der man-

nigfaltige Einsatz von Computern als einer besonderen Maschinenklasse semiotischer Maschinen, die sich fundamental von bisher üblichen energie- und stoffwandelnden Maschinen unterscheidet: Statt Energie oder Stoffe umzuwandeln werden Zeichen manipuliert; anstelle der Thermo- oder Elektrodynamik der Energiewandlung bzw. der Mechanik der Kraftübertragung bei herkömmlichen Maschinen tritt bei semiotischen Maschinen die Algorithmik von Berechnungsverfahren mit syntaktisch reduzierten, bedeutungslosen Quasi-Zeichen (›Daten‹) (vgl. III.; Brödner 2008, 2018). So verschleiert die Bezeichnung »Digitalisierung« gerade das Wesentliche: die Extraktion formalisierbarer Anteile geistiger Arbeit und deren Überführung in maschinell ausführbare Berechnungen, eine weitere »Entzauberung der Welt« (Weber 1919: 488).

2 »Künstliche Intelligenz«: Viel Lärm um nichts

Mit Blick auf die sog. »Künstliche Intelligenz« stellt sich zunächst die schlichte Frage: Was genau ist eigentlich ein »KI«-System? Darauf hält die einschlägige Literatur im wesentlichen zwei Gruppen unterschiedlicher Antworten parat. Die erste Gruppe begreift Computer als »KI«- bzw. »AI«-System, wenn die Lösung der Aufgaben, zu deren Bewältigung es geschaffen wird, natürliche Intelligenz erfordern:

- »… making a machine behave in ways that would be called intelligent if a human were so behaving.« (McCarthy 1955: 11);
- »AI is the part of computer science concerned with … systems that exhibit characteristics we associate with intelligence in human behaviour – understanding language, learning, reasoning, problem solving, and so on.« (Barr & Feigenbaum 1981);
- Computer 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: 9).

Eine zweite Gruppe von »KI«-Definitionen schreibt Computern eine gewisse eigenständige Handlungsträgerschaft (»agency«) zu:

- »AI is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment« (Nilsson 2010);
- AI research investigates »intelligent agents«, i.e. devices »that perceive their environment and take actions maximizing the chance of successfully achieving their goals« (Russell & Norvig 2009: 2);
- »Rational Agent«: »AI researchers use mostly the notion of rationality, which refers to the ability to choose the best action to take in order to achieve a certain goal, given certain criteria to be optimized and the available resources.« (EU High-Level Expert Group on AI 2018: 1f).

So schlicht und einfach die Frage erscheint, so verwirrend sind die Antworten, die zur Differenzierung zwischen »KI«-Systemen und ›gewöhnlichen‹ Computersystemen keinerlei brauchbare Einsicht liefern: Der ersten Definitionen zufolge erfordert doch auch die Lösung relativ einfacher Berechnungsaufgaben wie die Multiplikation großer Gleitkommazahlen, die Bestimmung der Nullstelle einer quadratischen Gleichung, das Sortieren einer Liste oder das Spiel der »Türme von Hanoi« beträchtliche Intelligenz. Sind folglich die dafür auf einem Computer verwendeten, sein Verhalten bestimmenden Berechnungsverfahren auch »KI«-Systeme? Nach dieser Definition wäre jedes auf einem Computer ausführbare Berechnungsverfahren ein »KI«-System, sie ist mithin unbrauchbar, die scheinbare *differentia specifica* unterscheidet nicht wirklich.

Bei der zweiten Gruppe von Definitionen werden Computersystemen die typischen Merkmale von Intentionalität und rationalem Handeln, die Wahl geeigneter Mittel zum erfolgreichen Erreichen von Zielen, einfach zugeschrieben. Dabei befolgen diese doch nur ein ihr Verhalten determinierendes Programm, dem bereits alle denkbaren Bedingungen eingeschrieben sind, unter denen zuvor festgelegte Ziele methodisch bestmöglich zu erreichen sind. Hier werden Machen und Gemachtes, die besonderen Tätigkeitsmerkmale des Entwerfens und Programmierens mit den Leistungen des Programms als deren Ergebnis verwechselt, ein krasser Kategorienfehler: Das Programm vergegenständlicht lediglich die Ergebnisse des lebendigen Arbeitsvermögens, der Intelligenz und Fähigkeiten seiner Schöpfer, vorgestellte Ziele unter angenommenen Bedingungen mit Mitteln der Logik und Verfahren der Berechnung bestmöglich zu verwirklichen.

Als Fazit bleibt mithin festzuhalten, dass aufgrund dieser Definitionen niemand weiß, schlimmer noch: niemand wissen kann, was ein »KI«-System eigentlich ist, paradoxerweise auch diejenigen nicht, die ständig davon reden – ein weiterer eklatanter Fall von »Technik und Wissenschaft als ›Ideologie‹« (Habermas 1968). Tatsächlich verkörpert jedes Computerprogramm allein die dem lebendigen Arbeitsvermögen geschuldete natürliche Intelligenz seiner Konstrukteure – ein Ergebnis, das freilich auch für jedes andere technische Artefakt gilt, vom Faustkeil bis zum Computer.

3 Zur Entmystifizierung der Funktionsweise von Computern

Zum Verständnis von Computersystemen sind offenbar genauere Einblicke in Aufbau und Funktionsweise binärer Schaltsysteme als deren materieller Grundlage sowie in die darauf operierenden elementaren logisch-mathematischen Funktionen nötig: Computer führen mittels der Schaltsysteme im streng mathematischen Sinne berechenbare Funktionen aus – zu nichts anderem sind sie befähigt (Kleene 1952; vgl. Brödner 1997, Kap. 2 & 4.4). Entworfen und eingesetzt werden sie entweder zwecks digitaler Steuerung materieller Prozesse (als »embed-

ded« bzw. »cyber-physical systems«) oder zur interaktiven Nutzung durch Menschen in deren sozialer Praxis als (»Informations-« bzw. »Organisationssysteme«).

Die Steuerung materieller Prozesse greift in Naturprozesse ein, basiert aber auf einer aus Einsicht in die physischen Wirkungsketten gewonnenen Prozessbeschreibung mittels Zeichen. Ein daraus entwickeltes hinreichend genaues mathematisches oder zumindest formal beschriebenes heuristisches Modell des jeweiligen Prozesses erlaubt dann, dafür ein Programm zu dessen digitaler Steuerung zu entwerfen, das aufgrund von relevanten Signalen aus dem Prozess zielführende Eingriffe in dessen Energie- oder Stoffströme auszulösen vermag, um dessen gewünschten Verlauf automatisch zu gewährleisten. Der interaktive Gebrauch von Computern greift hingegen in soziale Praktiken ein und beruht seinerseits auf dem verständigen Umgang mit Zeichen in diesen Praktiken.

In beiden Fällen dienen Computer mit ihren berechenbaren Funktionen der rein formalen Verarbeitung bedeutungsloser Zeichen(-körper). Unter einem Zeichen wird dabei die dreistellige Relation zwischen einem physischen Zeichen (-körper) (Repräsentamen R), dem damit bezeichneten Objekt (O) und der Bedeutung (Interpretant I) dieser Referenz in einem praktischen Handlungskontext verstanden: I – (R – O). »Ein Zeichen ist etwas, das für jemanden in einer bestimmten Hinsicht oder Fähigkeit für etwas steht.« (Peirce 1983).

In der Interaktion mit Computern werden von Benutzern Zeichen für Daten und damit operierende Funktionen, sog. »algorithmische Zeichen« (Nake 2001), eingegeben, die im jeweiligen sozialen Handlungskontext bestimmte Bedeutung tragen. Mit der Eingabe werden diese außen situativ sinnvoll interpretierbaren Zeichen auf bloße Signale als deren materiellen Verkörperungen (R) reduziert und mittels maschinell ausführbarer Anweisungen eines Programms, die das Berechnungsverfahren – den Algorithmus – formal beschreiben, verarbeitet (nach dem Modell der Turingmaschine; Turing 1936). Das mithin kausal determinierte Resultat R' dieser Signalverarbeitung kann dann bei Erscheinen an der Systemoberfläche wieder als Zeichen interpretiert werden. Beide Zeichenprozesse, der interne, auf programmgesteuerte, syntaktische Signalverarbeitung reduzierte wie die äußere sinngebende Interpretation, sind durch den gemeinsamen Zeichenkörper R auf der Benutzungsoberfläche fest gekoppelt: Anstelle der äußeren intentionalen Interpretation seitens der Benutzer ist die interne Verarbeitung der Signale (›Daten‹) durch die Programmanweisungen als kausalem Interpretanten determiniert. Das Ergebnis R' fällt mit dem auf diese Weise kausal berechneten Objekt zusammen. Eben die Kenntnis dieser Zusammenhänge ermöglicht dann außen eine sinnvolle Interpretation. So ist Interaktion mit Computern gekennzeichnet durch kausale Determination sinn- und kontextfreier Signal- bzw. Datenverarbeitung im Innern und durch sinngebende Interpretation der an der Oberfläche als Zeichen gedeuteten Signale oder Daten außerhalb. Der soziale Raum der Zeichenprozesse wird dabei nicht verlassen (Brödner 2008).

4 Mühen der Modellierung von Praxis und vergebliche Strategien ihrer Vermeidung

Gestaltung und Einsatz interaktiv genutzter Computer erfordern mithin die Modellierung und Formalisierung sozialer Praktiken, ein komplexer Vorgang, der auch leicht misslingen kann. Dabei werden der Praxis zugrunde liegende Zeichenprozesse zunächst in bestimmter Perspektive beschrieben durch partielle Explikation in Form begrifflich-propositionalen Wissens über deren Strukturen und Abläufe. Die Modellbildung – Kern des Software Engineering – erfordert natürliche Intelligenz und durchläuft folgende Schritte der Reduktion, Abstraktion und Formalisierung (Andelfinger 1997, Krämer 1988):

- Semiotisierung: Präzise Beschreibung einer sozialen Praxis mittels Zeichen liefert ein perspektivisch reduziertes Abbild von Wirklichkeit als Ergebnis gemeinsamer Reflexion und Kommunikation der Akteure (Sprachanalyse, ›Ontologie‹): → Anwendungsmodell.
- Formalisierung: Abstraktion von situations- und kontextgebundenen Bedeutungen und Reduktion auf sinnfreie Standardzeichen und -operationen:
  → formales Modell.
- Algorithmisierung: Überführung von Gegenständen und Abläufen des formalen Modells in auto-operational ausführbare Prozeduren in Form von Daten und berechenbaren Funktionen (Algorithmen):
  → Berechnungsmodell (als Grundlage der Programmierung).

Allerdings können so mittels Zeichen repräsentierte Denkvorgänge (»kognitive Funktionen«) stets nur partiell formalisiert, in berechenbare Funktionen überführt und als Algorithmus maschinell ausgeführt werden – auch Menschen rechnen formalisiert wie Maschinen, ihre Fähigkeiten sind aber nicht darauf beschränkt (daher gilt der Einsatz von Computern auch als »Maschinisierung von Kopfarbeit«; Nake 1992). Die Ausführung der berechenbaren Funktionen stellt einen »degenerierten«, auf eine dyadische Relation reduzierten Zeichenprozess ohne ›Fenster zur Welt‹ dar, dem der Bezug zu einem erlebten, leiblich erfahrenen oder gedachten Objekt, eben die ›Be-zeichnung‹ fehlt. Es ist nur eine »Quasi-Semiose«, die mit Signalen (›Daten‹) als auf Syntax reduzierten »Quasi-Zeichen« operiert (Nöth 2002). Deren Zustände werden per Programm rein physisch transformiert ohne Ansehen ihrer Bedeutung. Im Computersystem implementiert entstehen damit »auto-operationale Formen« (Floyd 2002) als Ausdruck abstrakter, formalisierter Operationen (vgl. oben »algorithmisches Zeichen«). Deren Sinn muss durch Aneignung seitens der Benutzer für wirksamen praktischen Gebrauch erst noch erschlossen werden. So werden Computersysteme durch Interpretation ihrer Funktionen im Handlungskontext einer sozialen Praxis wieder in einen – eben dadurch veränderten – Praxiszusammenhang gestellt.

In diesem Kontext gewinnt nun die allerdings meist ignorierte Unterscheidung zwischen Problem und Aufgabe hohe Bedeutung (Dörner 1983). Ein

Problem (griech. »*problematon*« = das zur Lösung Vorgelegte) liegt vor, wenn die Mittel zum Erreichen eines angestrebten Ziels noch unbekannt sind oder über das Ziel keine klaren Vorstellungen bestehen, wenn handelnde Personen also nicht wissen, wie sie ihr Ziel erreichen sollen: »Intelligenz ist das, was man einsetzt, wenn man nicht weiß, was man tun soll.« (J. Piaget). Gefordert sind dann Ideen für abduktives Schließen, d.h. die Bildung von erklärenden Hypothesen mittels Intuition, Analogie oder Kreativität (Peirce 1878). Hat sich die Hypothesenbildung bewährt, können daraus Verfahren zur methodischen Bewältigung von dem Problem entsprechenden Aufgaben gewonnen werden (Popper 1994).

Davon zu unterscheiden sind Aufgaben als gestellten geistigen Anforderungen eines bestimmten Typs, für deren Bewältigung Methoden oder Verfahren bereits bekannt sind. Aufgaben erfordern mithin lediglich den Einsatz bekannter Mittel auf bekannte Weise. Aufgaben sind gelöste Probleme, für deren Lösung Methoden oder Verfahren bereits verfügbar sind und nur abgerufen werden müssen. Ihre Lösung erfordert lediglich den routinierten Gebrauch dafür angeeigneter Methoden oder Verfahren (einschließlich der Beurteilung ihrer jeweiligen Eignung).

Die Modellierung einer bestimmten sozialen Praxis entspricht zunächst der Lösung eines Problems: Anfangs sind weder das Problem noch dessen Lösung hinreichend durchschaut und müssen im Zuge der Semiotisierung erst durch Analyse und Genese expliziten Wissens über die Praxis verstanden werden, um gesicherte Methoden der Bewältigung gewinnen zu können. Eben dadurch schrumpft das Problem zur Aufgabe, die dann durch Anwendung des Lösungsverfahrens bewältigt werden kann. In der Problemanalyse, der Wissensgenese, der Genese formalisierter Lösungsverfahren und der Beurteilung ihrer Eignung erweist sich die natürliche Intelligenz der Akteure, während die Leistung des Computersystems auf die Ausführung des programmierten Berechnungsmodells beschränkt ist, ggf. unter Berücksichtigung äußerer Bedingungen.

Mit der derzeit en vogue stehenden sog. »subsymbolischen KI« und der Nutzung von »Big Data« wird versucht, sich die Mühen von Problemanalyse, Modellierung, Formalisierung und Bestimmung eines adäquaten Berechnungsmodells zu ersparen. Stattdessen werden für eine breite Klasse von Aufgaben – darunter Aufgaben der Objekt-Klassifizierung bzw. - Identifizierung, der Clusterung von Objekten oder der Entscheidungsfindung – erfahrungsbasiert oder aufgrund theorielosen Probierens (»the end of theory«; Anderson 2008) einfach bestimmte mathematische Funktionen angenommen, deren Parameter dann noch aufgabenspezifisch bestimmt werden. Solche Funktionen können z.B. »künstliche neuronale Netze (KNN)« mit ihren Gewichten als Parametern, Polynome mit ihren Koeffizienten oder Entscheidungsbäume mit ihren Kantengewichten sein. Mittels längst bekannter Verfahren der Funktions-Approximation werden die Parameter möglichst gut an große Mengen vorgelegter Datenobjekte angepasst. Die so für die Bewältigung einer spezifischen Aufgabe »trainierten« Funktionen

werden dann auf neue Datenobjekte gleicher Art angewandt. Diese Verfahren sog. »maschinellen Lernens« haben aber nichts mit eigentlich reflexivem Lernen zu tun. Erfolg oder Misserfolg hängen von den zum »Training« benutzten Daten ab, deren Herkunft und Qualität aber meist nicht einschätzbar und hinsichtlich Repräsentativität und Verzerrungen (Bias) oft äußerst fragwürdig sind.

Der hohe Preis für ein solches Vorgehen ist, dass die Güte der Ergebnisse ungewiss bleibt und deren Interpretation unsicher ist – eine Art postmoderner Obskurantismus. Im Grunde kann man den berechneten Ergebnissen nur blind vertrauen, weil sich aktual nicht mehr nachvollziehen lässt, wie sie im einzelnen zustandekommen. Für die Mensch- Rechner-Interaktion und zweckorientiertes instrumentelles Handeln seitens der Nutzer hat diese Art undurchschaubaren Systemverhaltens allerdings höchst abträgliche Folgen, die reflexives Lernen behindern und großes Stresspotenzial aufweisen (Brödner 2020).

5 Erinnerung an prinzipielle Grenzen von Formalisierung

In diesem Zusammenhang ist ferner an prinzipielle Grenzen der Formalisierung von Zeichenprozessen und der Berechenbarkeit zu erinnern. Selbst die weitreichend formalisierbare Mathematik widersetzt sich ihrer vollständigen Algorithmisierung. Ausgerechnet im Zusammenhang mit den zu Beginn des 20. Jahrhunderts bestehenden großen Hoffnungen auf eine vollständige Formalisierung der bekannten Mathematik (›Hilbert-Programm‹) hat sich herausgestellt, dass es erwiesenermaßen unmöglich ist,

- einen Algorithmus anzugeben, der alle Sätze eines formalen Systems wie dem der Arithmetik 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 1936).

Bezeichnenderweise beruht der Beweis von Gödel im Kern darauf, dass er als erfahrener und kompetenter Mathematiker ein Verfahren gefunden hat, meta-mathematische Prädikate über das formale System der Arithmetik, z.B. das der Beweisbarkeit, als Formeln in diesem System selbst auszudrücken. Damit gelingt es ihm, ausschließlich mit Mitteln des formalen Systems eine Formel derart zu konstruieren, dass sie über einen durch ihn als wahr erkannten Satz aussagt, nicht beweisbar zu sein. Zur mathematischen Fähigkeit von Menschen gehört eben auch, dass sie über alles, was sie mithilfe dieser Fähigkeit operativ zu formalisieren vermögen, durch Nachdenken über die Formalisierung und intuitives, abduktives Schließen zu außer Reichweite der Formalisierung liegenden Einsichten gelangen können. Eben darin zeigt sich, was natürliche Intelligenz im Kern ausmacht (vgl. Brödner 1997; Kleene 1952).

Als rein formales Verfahren gibt die Abfolge von Operationen eines Algorithmus zwar Auskunft auf die Frage, wie genau etwas operativ abläuft; sie

beantwortet aber nicht die Frage nach deren Sinn oder Bedeutung, warum es so abläuft – eben deshalb gehört zur Software auch die Dokumentation mit derartigen Meta-Aussagen. Operationsfolgen sagen nichts über sich selbst aus, etwa ob sie korrekt oder gebrauchstauglich sind. Die logische Differenz zwischen Termen der operativen Ebene und Aussagen über dieselben auf der Metaebene lässt sich mit rein formalen Mitteln nicht überwinden. So ist etwa auch die Frage, ob ein Programm terminiert, formal per Algorithmus nicht entscheidbar.

6 Wider die »Verhexung unseres Verstandes«

Das aus der Entwicklungsgeschichte von Computertechnik und Computing Science gewonnene Verständnis der Genese und Funktionsweise von Computern als semiotischen Maschinen (vgl. z.B. Brödner 1997, Krämer 1988) erlaubt nun, die fundamentalen Unterschiede zwischen Menschen als lebendigen Organismen und Computersystemen als durch Menschen geschaffenen Artefakten aufzuzeigen (vgl. nachstehende Übersicht). Die ständige Rede von »intelligenten« oder gar »autonomen« Computersystemen entpuppt sich dabei als folgenreicher Etikettenschwindel. Einmal mehr bedarf es der Philosophie als »Kampf gegen die Verhexung unseres Verstandes durch die Mittel unserer Sprache« (Wittgenstein 1984: PU 109).

Übersicht: Ontologische Differenz zwischen Mensch und Computer

Mensch *(lebendiger Organismus)*	***Computer*** *(semiotische Maschine)*
Sich mittels Autopoiese in Stoffwechsel und Kommunikation *selber machend*.	Wissensbasiert für *bestimmte Zwecke gemacht* (methodisch konstruiert).
Autonom (nach selbstbestimmten Regeln).	*Automatisch* (programmiert auto-operational).
Handelt *intentional* (kontingent), ist *sprachbegabt*, *reflexiv* lernfähig.	Verhält sich *kausal determiniert*; ggf. algorithmisch gesteuert *Umwelt-adaptiv* (mittels Funktions-Approximation).
Lebendiges *Arbeitsvermögen*: *Können* (implizites Wissen, Erfahrung, situierte Urteilskraft, Handlungskompetenz), *verausgabt* & *reproduziert* sich im Gebrauch.	Algorithmisch *determiniertes Verhalten:* Setzt *Formalisierung* von Zeichenprozessen voraus, muss für Praxis *angeeignet* & *organisatorisch eingebettet* werden.

Die irreführend auf Computer angewandte Metaphorik, als ob diese wie Menschen »intentional eingestellt« und zweckorientiert handlungsfähig wären – »intelligent«, »autonom«, »selbstlernend«, »selbstorganisierend« oder gar »selbst--heilend« etc. – ignoriert nach Denkmustern des Funktionalismus die fundamentalen Unterschiede in extrem reduktionistischer Weise. Indem mentale als bloß funktionale Zustände begriffen werden, als unabhängig vom materiellen Medium ihrer Realisierung, wird zum einen kompetentes Handeln von Menschen auf

algorithmisch gesteuertes Verhalten von Maschinen reduziert; zum anderen entstehen eben dadurch Illusionen über dessen Zustandekommen und tatsächliche Leistungsfähigkeit. Das führt im Ergebnis zu einer verbreiteten Selbsttäuschung, wie sie in »KI«-Diskursen oft zum Ausdruck kommt (Brödner 2018). In den Illusionen und Wahnvorstellungen vom vermeintlichen Eigenleben der Maschinen äußert sich deren Fetischcharakter, eben die »Macht der Machwerke über die Machenden« (Haug 2005: 162).

7 Fazit: Auf realistische Entwicklungsperspektiven kommt es an

Die Zunft der dieser Selbsttäuschung anheim gefallenen »KI«-Adepten hat derzeit wieder großen Zulauf; sie ist indes großenteils in Heilsverkünder und Apokalyptiker gespalten.

Beiden gemeinsam ist ihre Immunität gegen Tatsachen und die Unkenntnis der tatsächlichen Berechnungsvorgänge in Computersystemen wie auch der Methoden ihrer Entstehung. Apokalyptiker befürchten irrigerweise, dass Computersysteme, obgleich von Menschen als semiotische Maschinen geschaffen, schon bald die Menschheit insgesamt an Intelligenz überflügeln und unkontolliert die Macht über sie übernehmen könnten; tatsächlich geht es aber wie stets um die Macht von Menschen über andere Menschen (vermittelt über die Maschinen). Als die wahre, viel furchterregendere Horrorvision erscheint hingegen die gesellschaftliche Dominanz von Menschen, die sich selbst auf algorithmisch gesteuertes Verhalten reduziert begreifen, folglich wie Maschinen denken und handeln.

Dabei kommt am Ende alles auf die Entwicklungsperspektive künftiger Computertechnik an, auf die Art und Weise, wie deren Funktionsweise begriffen und zu welchen Zwecken sie angeeignet und eingesetzt wird: Statt weiterhin fragwürdige »KI«-Konzepte zur Automatisierung kognitiver Wissensarbeit zu verfolgen, sollte die Gestaltung guter Arbeit und vor allem die Entwicklung genuin menschlicher Fähigkeiten im Fokus stehen. Um Produktivität, Innovation und gesellschaftlichen Wohlstand auch künftig zu ermöglichen, muss sich, wie viele arbeitswissenschaftliche Erkenntnisse der letzten Dekaden lehren, die Gestaltung soziotechnischer Systeme an Erfordernissen menschlichen Handelns, dabei v.a. an Möglichkeiten der Entfaltung lebendigen Arbeitsvermögens, orientieren. Dafür gilt es nützliche und nutzbare Computerartefakte zu schaffen, die dann auch auf gesellschaftlich sinnvolle Weise als instrumentelles Medium der Kooperation genutzt werden können. Genau diese zentralen Gestaltungsaspekte werden mittels der vermeintlich ›Modernität‹ vortäuschenden Benennung »Digitalisierung« ausgeblendet und durch Bemühungen um »KI« hintertrieben; stattdessen wird damit einem naiven Technikdeterminismus gehuldigt.

↑Inhalt↑

8 Literatur

Andelfinger, U. (1997): Diskursive Anforderungsanalyse. Ein Beitrag zum Reduktionsproblem bei Systementwicklungen in der Informatik, Frankfurt/M: Peter Lang

Anderson, C. (2008): The End of Theory, Wired 23.06.08

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

Barr, A. & Feigenbaum, E.A. (1981): The Handbook of Artificial Intelligence, Stanford (CA): HeurisTech Press

Brödner, P. (2020): Paradoxien der Koaktion von Experten und adaptiven Systemen, in: P. Brödner & K. Fuchs-Kittowski (Hg.): Zukunft der Arbeit – soziotechnische Gestaltung der Arbeitswelt im Zeichen von »Digitalisierung« und »Künstlicher Intelligenz«, Abhandlungen der Leibniz-Sozietät der Wissenschaften Band 67, Berlin: trafo Wissenschaftsverlag, 143-159

Brödner, P. (2018): Coping with Descartes' Error in Information Systems, AI & Society Journal of Knowledge, Culture and Communication 2018 (online first)

Brödner, P. (2008): Das Elend computerunterstützter Organisationen, in: Gumm, D.; Janneck, M.; Langer, R.; Simon , E. J. (Hg.): Mensch –Technik – Ärger? Zur Beherrschbarkeit soziotechnischer Dynamik aus transdisziplinärer Sicht, Münster: Lit-Verlag, 39-60

Brödner, P. (1997): Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen, Berlin: edition sigma

Dörner, D. (1983): Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität. Bern: Huber EU High-Level Expert Group on Artificial Intelligence (2018:): A Definition of AI: Main Capabilities and Scientific Disciplines, Brussels: European Commission

Floyd, C. (2002): Developing and Embedding Autooperational Form, in: Dittrich, Y.; Floyd, C.;

Klischewski, R. (Eds.): Social Thinking – Software Practice, Cambridge (MA): MIT Press, 5-28

Gödel, K. (1931): Über formal unentscheidbare Sätze der principia mathematica und verwandter Systeme I, Monatshefte für Mathematik und Physik 38 (1), 173-198

Habermas, J. (1968): Technik und Wissenschaft als ›Ideologie‹, Frankfurt/M: Suhrkamp

Haug, W. F. (2005): Vorlesungen zur Einführung ins »Kapital«, Hamburg: Argument

Hausdorff, F. (1897): Sant' Ilario, Gedanken aus der Landschaft Zarathustras (unter dem Pseudonym Paul Mongré), zitiert in Semiosis 19, Int. Zeitschrift für Semiotik und Ästhetik 5 (3), 69

Kleene, S.C. (1952): Introduction to Metamathematics, Amsterdam: North Holland

Krämer, S. (1988): Symbolische Maschinen. Die Idee der Formalisierung in geschichtlichem Abriss, Darmstadt: Wissenschaftliche Buchgesellschaft

McCarthy, J. (1955): A Proposal for the Summer Research Project on Artificial Intelligence, http://www-formal.stanford.edu/jmc/history/dartmouth.pdf

McCarthy, J. (1960): Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I, CACM 3 (4), 184-195

Nake, F. (2001): Das algorithmische Zeichen, in: Bauknecht, W.; Brauer, W.; Mück, T. (Hg.): Informatik 2001. Tagungsband der GI/OCG Jahrestagung, 736-742

Nake, F. (1992): Informatik und die Maschinisierung von Kopfarbeit, in: Wolfgang Coy et al. (Hg.): Sichtweisen der Informatik, Brauchschweig Wiesbaden: Vieweg, 181-201

Nilsson, N.J. (2010): The Quest for Artificial Intelligence: A History of Ideas and Achievements, Cambridge (UK): Cambridge University Press

Nöth, W. (2002): Semiotic Machines, Cybernetics and Human Knowing 9 (1), 5-22

Peirce, C.S. (1983): Phänomen und Logik der Zeichen, Frankfurt/M: Suhrkamp

Peirce, C.S. (1878): Deduction, Induction, and Hypothesis, in: Collected Papers Vol. 2, ed. by C. Hartshorne, & P. Weiss, P., Cambridge (MA): Harvard University Press (1931-35)

Popper, K.R. (1994): Alles Leben ist Problemlösen. Über Erkenntnis, Geschichte und Politik, München. Piper

Pylyshyn, Z. (1984): Computation and Cognition. Towards a Foundation of Cognitive Science, Cambridge, MA: MIT Press

Turing, A.M. (1950): Computing Machinery and Intelligence, Mind 49, 433-460

Turing, A.M. (1936): On Computable Numbers. With an Application to the Entscheidungsproblem, J. of Symbolic Logic, 230-265

Weber, M. (1919): Wissenschaft als Beruf, in: D. Kaesler (Hg.): Max Weber Schriften 1894 - 1922, Stuttgart: Alfred Kröner Verlag 2002

Wittgenstein, L. (1984): Philosophische Untersuchungen, Werkausgabe Bd. 1, Frankfurt/M: Suhrkamp

↑Inhalt↑

Paradoxien der Ko-Aktion von Experten und adaptiven Systemen

1 Einführung: Der Rummel um ›künstliche Intelligenz‹

In den letzten Jahren ist im öffentlichen Raum wie in der Computertechnik unter dem Etikett ›Industrie 4.0‹ wieder viel von vermeintlicher ›künstlicher Intelligenz‹ (KI) in Gestalt ›lernender‹ oder ›autonomer Systeme‹ die Rede. Nach der Aufregung um die ›Kybernetik‹ in den 1950er Jahren und dem Gespenst der ›menschenleeren Fabrik‹ in den 1980gern ist es nun schon die dritte Welle einer angekündigten Umwälzung, die freilich bei seriöser Prüfung der Konzepte abermals an der Realität zu scheitern droht (Brödner 2017). Lagen wesentliche Gründe des Scheiterns bei der ›Kybernetik‹ theoretisch in der Begriffsverwirrung um ›Information‹ und praktisch in der Unzulänglichkeit der Rechner-Hardware, bei der ›symbolischen‹ oder ›wissensbasierten KI‹ der 1980er Jahre in der prinzipiell begrenzten, meist unzureichenden Explizierbarkeit praktischen Könnens in propositionales, formal manipulierbares Wissen, so werden derzeit in alternativ verfolgten datengetriebenen sog. ›subsymbolischen‹ Ansätzen v. a. methodische Schwierigkeiten offenkundiger.

Erst jüngst hat die EU-Kommission (2020: 1) in ihrem Weißbuch zur ›KI‹ festgestellt: »Die Künstliche Intelligenz entwickelt sich schnell. Sie wird unser Leben verändern, indem sie die Gesundheitsfürsorge verbessert (z.B. durch präzisere Diagnostik und bessere Prävention von Krankheiten), die Effizienz der Landwirtschaft erhöht, zum Klimaschutz und zur Anpassung an den Klimawandel beiträgt, die Effizienz von Produktionsanlagen durch vorausschauende Wartung steigert, …«. Fast zeitgleich verweist ein Fachkundiger auf sich immer deutlicher abzeichnende Schwächen der neuen ›KI‹: »To the contrary, deep learning techniques thus far have proven to be data hungry, shallow, brittle, and limited in their ability to generalize« (Marcus 2020: 4).

Seit jeher sind die vollmundigen Versprechen über Leistungen der ›KI‹ weit größer als die tatsächlich erreichten Ergebnisse. Theoretischer und konzeptioneller Fortschritt bleibt eng begrenzt, während reale Verbesserungen hauptsächlich auf zwischenzeitlich exponentiell gesteigerter Leistung der Hardware beruhen. Gleichwohl lösen sie im öffentlichen Raum jedesmal große Befürchtungen technisch bedingter Arbeitslosigkeit aus (›Computer als Jobkiller‹). So haben fast drei Viertel (73 %) der Menschen in Deutschland einer aktuellen Umfrage zufolge Angst vor einem Verlust ihres Arbeitsplatzes durch den technologischen Wandel

(vgl. heise online 26.01.2020). Der lässt sich aber empirisch nicht nachweisen. Noch immer gilt das Produktivitäts-Paradoxon der Computertechnik, die sich trotz jahrzehntelanger massiver Investitionen so gut wie gar nicht verstärkend auf die gesamtwirtschaftliche Arbeitsproduktivität auswirkt (wohl aber Strukturwandel etwa bei Tätigkeiten und Berufen verursacht; vgl. Weber et al. 2016). Entgegen landläufiger Meinung ist seit der Jahrtausendwende die Arbeitsproduktivität in entwickelten Gesellschaften sogar auf das Niveau vor der industriellen Revolution gesunken (OECD productivity statistics).

Vor diesem Hintergrund hochgradig umstrittener künftiger Entwicklung der Computertechnik lohnt es sich, mögliche Tendenzen und Schwierigkeiten des Zusammenwirkens von Menschen mit fortschrittlichen Computersystemen genauer zu betrachten und daraus Schlüsse für die soziotechnische Gestaltung von Arbeit zu ziehen. Dazu wird im Folgenden zunächst Einsicht in die tatsächliche Genese und Funktionsweise fortschrittlicher Computersysteme genommen, um angesichts fortdauernder Notwendigkeit der Interaktion von Experten mit solchen Systemen näher zu bestimmen, welche neuen oder besonderen Herausforderungen sich dabei ggf. ergeben. Eingedenk des nicht mehr durchschauten Black-Box-Verhaltens der immer komplexer werdenden Systeme und der Unsicherheit ihrer Resultate ist eher mit besonderen psychischen Belastungen und Stressreaktionen zu rechnen. Mittels eines avancierten, empirisch erprobten relationalen Modells der Stressgenese werden diese Besonderheiten analysiert, um auf dieser Grundlage Schlussfolgerungen für die soziotechnische Gestaltung der Systeme zu ziehen.

2 Die Mär ›autonomer Agenten‹ und die meist übersehenen Risiken der Adaptivität

Auf Selbsttäuschung hinauslaufende Mystifizierungen oder Wahnvorstellungen sind in der Kulturgeschichte keine Seltenheit, auch nicht in Wissenschaft und Technik. Meist gehen sie – von vorherrschenden Interessen geleitet – auf irreführende Begriffe und Bezeichnungen zurück, werden durch Fachsprache scheinbar wissenschaftlich verbrämt oder durch Äußerungen wissenschaftlicher Autoritäten beglaubigt. So wird mit der ständigen Rede von angeblich ›autonomen Agenten‹ oder ›lernenden‹ Systemen sowie von Algorithmen, die »Spiele spielen«, »Bilder erkennen« oder »Sprache übersetzen« gerade das verborgen, was sie tatsächlich ausmacht, dass sie erstens zur Gänze fremdbestimmten Anweisungen folgen, mithin wohl determinierte Operationen ausführen, und dass sie zweitens von ihren Konstrukteuren für jeweils genau bestimmte Zwecke geschaffen wurden. In aller Regel treten die in dieser Sprache zum Ausdruck kommenden Einbildungen oder Selbsttäuschungen paarweise in einer apokalyptischen und einer heilsversprechenden Variante in Erscheinung. Deutlich zeigt sich ein derartiger wahnhafter Zwilling beispielsweise in dem – nach dreißigjähriger Pause –

erneut aufgeflammten Diskurs um vermeintliche ›künstliche Intelligenz‹. In maßloser Überschätzung der tatsächlichen Leistungsfähigkeit der darunter subsumierten Systeme warnen einerseits etwa viele sich dazu berufen fühlende Wissenschaftler in einem offenen Brief vor den Gefahren außer Kontrolle geratender »intelligenter Maschinen« (heise online 14.01.2015; Russell et al. 2015), während andererseits Lovelock (2019) mit dem »Novazän« die eher heilsbringende Variante einer maschinellen »Hyperintelligenz« propagiert.

Beide Varianten sind auf Vergleiche äußerlichen Verhaltens von Menschen und Maschinen bei bestimmten Aufgaben fixiert (so schon McCarthy 1955). Gestützt auf längst widerlegte Hypothesen des Funktionalismus (Putnam 1960, Fodor 1968), wird so intelligentes Handeln von Menschen als lebendigen, zu Bewusstsein und Reflexion fähigen Organismen unter der Hand auf die Ausführung berechenbarer Funktionen durch elektronische Schaltsysteme reduziert. Statt sich, durch diesen unzulässigen Reduktionismus gegen Tatsachen und methodische Einwände immunisiert, mittels irreführender anthropomorphisierender Metaphern in phantastische ›Fiction‹ hineinzusteigern, wäre es weit ergiebiger und dem Erkenntnisfortschritt zuträglicher, methodisch gesicherte ›Science‹ zu betreiben und sich der tatsächlichen Genese und Funktionsweise von Computersystemen zu vergewissern, denen fälschlich ›autonomes‹, ›intelligentes‹ oder ›lernendes‹ Verhalten zugeschrieben wird. Dann würde schnell offenkundig, dass das Verhalten dieser Systeme – wie das konventioneller Computersysteme auch – ungeachtet ihrer Komplexität ausschließlich in der Ausführung Turing-berechenbarer Funktionen auf binären Schaltsystemen besteht; zu nichts anderem sind Computer ja auch fähig. Ihre oftmals erstaunlichen Leistungen verdanken sie daher allein dem Können, der Erfahrung und der methodischen Kompetenz, mit denen ihre Konstrukteure die jeweils dem Einsatz zugrunde liegenden Zwecke und Probleme analytisch zu durchdringen und zu deren Bewältigung mittels oft sehr aufwendiger Formalisierung und Abstraktion wirksame Lösungsverfahren in Form ausführbarer Berechnungsmodelle zu entwickeln vermögen (Brödner 2019a, 2019b). So bleiben denn auch die realen Fortschritte entgegen landläufiger Auffassung eher bescheiden; meist beruhen berichtete Fortschritte angesichts geringer konzeptioneller Verbesserungen auf alt bekannten Verfahren, ausgeführt auf exponentiell in der Leistung gesteigerter Hardware.

Seit jeher erfordert der praktische Einsatz von Computersystemen eine sorgfältige und aufwendige Modellierung zeichenbasierter sozialer Praktiken von kognitiver bzw. Wissensarbeit: Mittels Computersystemen kann allenfalls »Kopfarbeit maschinisiert« werden (Nake 1992), wenigstens partiell. Dazu müssen der Problemlösung zugrunde liegende Zeichenprozesse (»Semiosen«, Peirce 1983) in bestimmter Perspektive sachgerecht modelliert werden durch partielle Explikation sozialer Praktiken in Gestalt begrifflichen Wissens über deren Strukturen und Abläufe. Dabei unterliegt die Modellbildung – Kern des Software Engineering –

stets Interessen und Machteinflüssen beteiligter Akteure und durchläuft folgende Schritte der Abstraktion und Formalisierung (Andelfinger 1997):

- Semiotisierung: Die zunehmend präzisierte Beschreibung der betrachteten sozialen Praxis mittels Zeichen liefert ein perspektivisch reduziertes Abbild derselben als Ergebnis gemeinsamer Reflexion und Kommunikation der Akteure (Sprachanalyse, →›Ontologie‹). Ergebnis ist ein im wesentlichen sprachlich artikuliertes *Anwendungsmodell*.
- Formalisierung: Dessen Formalisierung durch Abstraktion von situations- und kontextgebundenen Interpretationen mittels Verwendung standardisierter Zeichen und Operationen (funktionale Spezifikation) liefert ein *formales Modell*.
- Algorithmisierung: Die Überführung von Objekten und Abläufen des formalen Modells in auto-operational ausführbare Prozeduren in Form von Daten und berechenbaren Funktionen (Algorithmen) liefert schließlich das *Berechnungsmodell*.

Die berechenbaren Funktionen operieren auf Daten als auf syntaktische Aspekte reduzierten »Quasi-Zeichen« (Nöth 2002), deren binäre Verkörperungen (»Repräsentamen«, Peirce 1983) sie imperativ verändern ohne Ansehen ihrer Bedeutung. Im Computersystem implementiert bilden sie »auto-operationale Formen« als Ausdruck abstrakter, formalisierter Handlungen (Floyd 2002). Deren Sinn muss durch Aneignung für den praktisch wirksamen Gebrauch erst wieder erschlossen werden. So werden die Computersysteme durch Interpretation ihrer Funktionen im Handlungskontext der Arbeitspraxis, wieder in einen – freilich eben dadurch veränderten – Praxiszusammenhang gestellt.

In Verkennung der unüberwindlichen logischen Differenz zwischen zweckmäßig gemachten Artefakten und den sie intentional und einsichtsvoll Machenden wird seitens der Wahnvorstellungen ›künstlicher Intelligenz‹ beharrlich ignoriert, dass Computersysteme gleich welcher Komplexität, wie übrigens klassische Maschinen der Energie- und Stoffumwandlung schon immer, stets lediglich das zur Problemlösung erforderliche explizite methodische und funktionale Wissen vergegenständlichen, das sich der natürlichen analytischen Intelligenz ihrer Konstrukteure verdankt. Den zwecks Maschinisierung körperlicher Arbeit geschaffenen mechanischen Funktionen der Kraftübertragung als dem Kern maschineller Energie- und Stoffumwandlung entsprechen dann bei Computern als semiotischen Maschinen die algorithmischen Funktionen für Datenverarbeitung, - speicherung und -transfer zwecks Maschinisierung kognitiver Wissensarbeit. Als gemachte, programmbefolgende Artefakte sind sie folglich selbst weder ›autonom‹ noch ›intelligent‹ und bleiben bei vielen Aufgaben auf die Mitwirkung von Wissensarbeitern angewiesen, die sich dafür der Systemfunktionen bemächtigen müssen. Gelegentlich lassen sich jedoch zur Bewältigung bestimmter, genau begrenzter Aufgaben selbsttätig ausgeführte maschinelle Funktionen finden; sie bilden dann aufgabenspezifische, menschlicher Leistung oft weitaus überlegene Automaten,

nicht aber ›autonome‹ (d.h. nach selbstbestimmten Regeln operierende) Systeme, denn tatsächlich ist ihr Verhalten durch ihre Konstrukteure zur Gänze fremdbestimmt.

Auch das Computersystemen angedichtete ›maschinelle Lernen‹ ist mit menschlichem Lernen in keiner Weise vergleichbar, suggeriert gleichwohl Ähnlichkeit (und wird meist auch so verstanden). Tatsächlich geht es jedoch bei den mathematischen Verfahren sog. ›maschinellen Lernens‹ gleich welcher Art lediglich um bloße Funktions-Approximation an gegebene Datenobjekte: Dabei häufig zugrunde gelegte Berechnungsmodelle oder -verfahren, beispielsweise ›künstliche neuronale Netze‹ (KNN; vgl. Kriesel 2005), sog. ›Vector Support Machines‹ (vgl. Burges 1998), Entscheidungsbaum Verfahren (vgl. Nilsson 1998), Verfahren zur ›Cluster-Analyse‹ (vgl. McCay 2003) oder lineare Regressionen, werden mit Hilfe einer Kosten- oder Verlustfunktion als Gütemaß während einer ›Trainingsphase‹ an eine meist große Menge gegebener Datenobjekte optimal angepasst, um damit anschließend weitere von außen aufgenommene Daten bestmöglich klassifizieren oder zuordnen zu können. Die mittels dieser Verfahren durchgeführten Berechnungen liefern jedoch allesamt nur wahrscheinlich zutreffende, daher grundsätzlich unsichere Ergebnisse. Aufgrund dieser Möglichkeiten datengetriebener Verhaltensanpassung werden solche Computersysteme auch korrekterweise als *adaptive Systeme* bezeichnet.

Gleichwohl werden adaptive Systeme, die auf bestimmte Umweltbedingungen programmgesteuert automatisch »rational« und zielverfolgend zu reagieren vermögen, in der Literatur gleich wieder als »intelligente Agenten« mystifiziert (Russell & Norvig 2009). Tatsächlich sind sie lediglich mit einer Mensch-Maschine-Schnittstelle ausgestattet, können über Sensoren Signale aus ihrer Umgebung aufnehmen, damit ›passende‹ Reaktionen berechnen und über Aktoren auf diese wiederum einwirken sowie Daten mit anderen Systemen austauschen (vgl. Abb. 1).

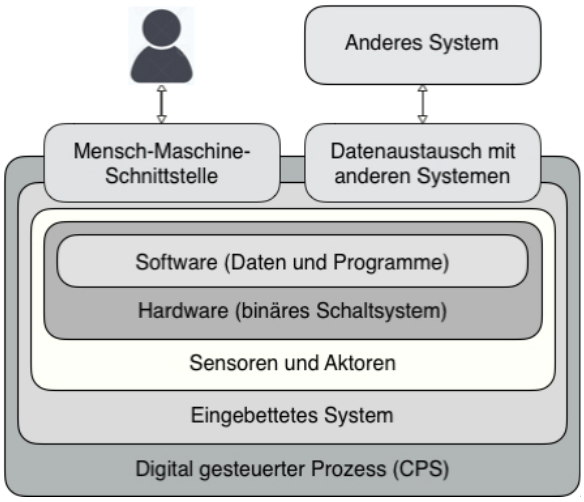


***Abb. 1**: »Software-Agent« als adaptives System (eigene Darstellung, angelehnt an Broy 2010)*

Diese adaptiven, sich algorithmisch gesteuert gegebenen Daten aus der Umwelt anpassenden Systeme bilden nun tatsächlich eine besondere Klasse von Computersystemen: Gegenüber nicht-adaptiven Systemen weisen sie u.a. eine höhere Stufe der Undurchschaubarkeit und Unsicherheit auf; ihre Funktionsweise hängt nun nicht mehr allein vom implementierten Programm(-system), sondern auch noch von den jeweils für die Funktions-Approximation verwendeten Datenobjekten ab. Ihre Leistungsfähigkeit steht und fällt daher mit der Güte der zur Anpassung verwendeten Daten. Deren Qualität, etwa hinsichtlich Erhebungsmethodik, Representativität oder Validität, ist aber meist fragwürdig, durch unbekannte Verzerrungen kontaminiert und daher auch kaum im Vorhinein einschätzbar.

Oftmals ist die Funktionalität einzelner adaptiver Systeme oder »Software-Agenten« (Bradshaw & Hutchinson 1997) auf die selbsttätige Bewältigung relativ begrenzter Aufgaben beschränkt. Sollen dann auch komplexere Aufgaben, etwa in einer dynamischen physischen Umwelt bewältigt werden, lassen sich solche begrenzt automatischen Software-Agenten derart einrichten und miteinander zu einem Netzwerk verbinden, dass sie ihr Verhalten mittels Datenaustausch zu koordinieren vermögen, um so komplexere Aufgaben durch Interaktion in einem Multi-Agenten-System (MAS) gemeinsam zu erledigen (Wooldridge 2002).

Auch wenn einzelne im Systemverbund interagierende ›Agenten‹ relativ einfache Programme befolgen und gut durchschaubares Verhalten aufweisen, zeigt das MAS insgesamt ein zwar vollständig determiniertes, aber hoch komplexes und analytisch von außen nicht bestimmbares Verhalten. Formal lassen sich MAS als sog. »nicht- triviale Maschinen« (v. Foerster 1993) beschreiben, deren Ausgabedaten nicht nur von den Eingabedaten, sondern gemäß einer Zustandsfunktion auch von veränderlichen inneren Zuständen abhängen, die auf vielfältige Weise die Interaktion der Agenten und deren adaptives Verhalten zum Ausdruck bringen (vgl. Abb. 2).

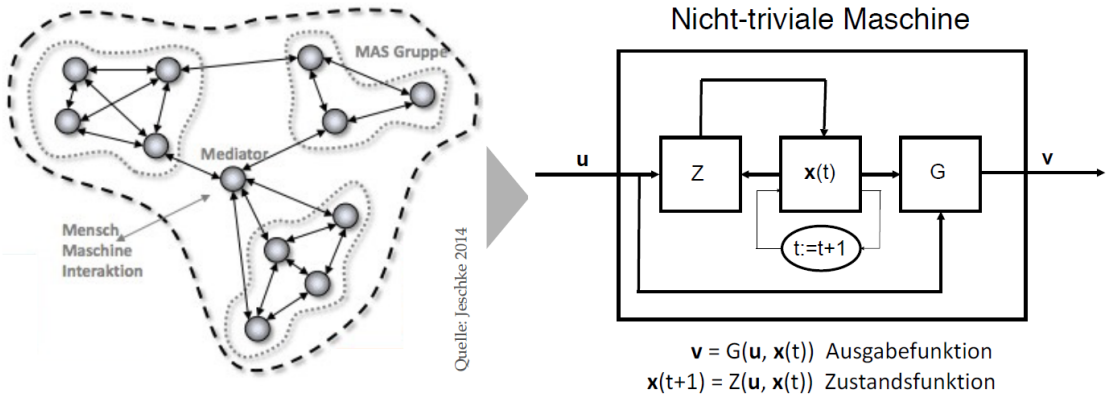


Abb. 2: *Nicht-triviale Maschine (v. Foerster 1993) als Modell von Multiagentensystemen (eigene Darstellung)*

Infolgedessen ist das Verhalten von MAS in hohem Maße von der jeweiligen Vorgeschichte abhängig, analytisch nicht mehr bestimmbar und mithin auch nicht vorhersehbar. Ihr situatives Verhalten ist für Beobachter nicht mehr nachzuvollziehen. Wenn sich solche Systeme in vergleichbaren Situationen jeweils anders und unerwartet verhalten, bereitet das ihren Benutzern beträchtliche Schwierigkeiten, sich diese für praktisch wirksamen Gebrauch anzueignen und mit ihnen zweckmäßig und zielgerichtet zu interagieren. Damit würde eklatant gegen eine der Grundregeln der Mensch-Maschine-Interaktion verstoßen, gegen die Forderung nach erwartungskonformem Verhalten (vgl. EN ISO 9241-11: Anforderungen an die Gebrauchstauglichkeit). Damit wird ein Zusammenwirken von Wissensarbeitern und adaptiven Systemen, das in Anbetracht ihrer i.a. begrenzten Leistungen auch in der absehbaren Zukunft erforderlich bleibt, auf eine neue, bislang kaum beachtete Stufe gehoben.

So wirft der instrumentelle Gebrauch adaptiver Systeme eine Reihe gravierende Probleme sowohl epistemischer wie ethischer Natur auf, macht ihn im Grunde nahezu unmöglich: Das Verhalten von ›KI‹-Algorithmen – hier v.a. künstliche neuronale Netze (KNN) und Verfahren induktiver Statistik – ist selbst für ihre Entwickler aktuell weder im einzelnen durchschaubar (»inconclusive evidence«) noch im Nachhinein erklärbar (»inscrutable evidence«). Sie produzieren nur wahrscheinliche, daher stets unsichere Ergebnisse, deren Korrektheit und Validität nur schwer zu beurteilen sind. KNN sind zudem sehr störanfällig und leicht auszutricksen. Die Ergebnisse, die sie liefern, sind in hohem Maße von der Qualität der Eingabedaten abhängig, die aber meist ebenfalls unbekannt oder nur schwer einschätzbar ist (»misguided evidence«). Zugleich werden auf Seiten der Nutzer stets aufs Neue überzogene Erwartungen an die »Handlungsfähigkeit« der Systeme geschürt, mithin gar ihre Wahrnehmung der Wirklichkeit verändert (»transformative effects«; vgl. Mittelstadt et al. 2016: 4f; Brödner 2019b).

Finden sich Wissensarbeiter als Nutzer solcher Systeme mit derartigen Widersprüchen konfrontiert, sind sie insbesondere hohem Erwartungsdruck erfolgreicher Bewältigung ihrer Aufgaben unterworfen bei gleichzeitigem Verlust der Kontrolle über Arbeitsmittel mit undurchschaubarem und ungewissem Verhalten, so werden sie dadurch dauerhaft psychischen Belastungen ausgesetzt (so Norman bereits 1994 im Hinblick auf ›Agenten‹ als frühen adaptiven Systemen; vgl. auch Bradshaw et al. 2011; Kabel et al. 2001). Diese neuartige Belastungssituation besteht zumindest, solange diese Systeme nicht in die Lage versetzt werden können, ihr eigenes Verhalten situationsspezifisch in erforderlichen Einzelheiten zu erklären, was derzeit – obgleich häufig gefordert – bislang jedoch in weiter Ferne liegt (vgl. etwa Kindermans et al. 2017, Park et al. 2017). Auch lässt sich dann von Interaktion im Sinne gewohnten instrumentellen Gebrauchs von Computersystemen nicht mehr sprechen; vielmehr handelt es sich infolge des veränderlichen Verhaltens um *Ko-Aktion* von Menschen und adaptiven Systemen

(Hubig 2019), die ihre Betreiber und Nutzer vor besondere Herausforderungen stellen.

So werden mit dem Einsatz adaptiver Systemen auch Fragen nach der ethischen Verantwortbarkeit aufgeworfen: Dürfen derart undurchschaubare und störanfällige Artefakte überhaupt eingesetzt werden, da man ihre Resultate nicht zuverlässig beurteilen kann, sich blind auf sie verlassen muss? Wer ist gegebenenfalls für eingetretene Schäden verantwortlich – die Entwickler?, die Betreiber? oder gar die Nutzer? – und wie werden daraus entstehende Haftungsansprüche geregelt? Bisher dazu getroffene oder sich abzeichnende Regelungen sind noch unzureichend und unbefriedigend.

Auch hier gelten seit langem diagnostizierte »Ironien der Automatisierung« (Bainbridge 1983) zugespitzt weiter: Von besonderer Bedeutung ist dabei die Ironie, dass ausgerechnet die im Stör-oder Versagensfall unverzichtbare menschliche Handlungskompetenz mangels Aneignungsmöglichkeiten gar nicht ausreichend entwickelt wird bzw. mangels hinreichender Übung und Erfahrung bei automatischem Normalbetrieb schwindet. Es fehlen organisatorische Konzepte, wie dieser Art ›erlernter Inkompetenz‹ entgegengewirkt werden kann. In Schadensfällen wird die Ursache meist dichotomisch in ›menschlichem Versagen‹ oder in einer technischen Störung gesucht und dabei die wahre, in fehlgeleiteter soziotechnischer Arbeitsgestaltung liegende Ursache ignoriert. Die tatsächlich jedoch systemisch bedingte mangelnde Kompetenz, in Verbindung mit der durch Untätigkeit oder Ablenkung geschwächten Vigilanz, führt dann bei plötzlich notwendigen Eingriffen zu beträchtlichen Problemen der Bewältigung, etwa zu fehlerhaften Diagnosen und falschen oder riskanten Handlungen.

Beispiele und empirische Erkenntnisse zu diesen letzteren Herausforderungen gibt es aus der Forschung über Leitwartentätigkeiten, Flugführung oder automatisiertes Fahren zuhauf, gleichwohl wurden bislang wenig weiterführende Konsequenzen gezogen. Daher ist infolge dieser noch weitgehend ungelösten Schwierigkeiten im praktischen Einsatz und Gebrauch adaptiver Systeme in naher Zukunft mit beträchtlichen Verzögerungen der Entwicklung zu rechnen (vgl. Bainbridge 1983, Baxter et al. 2012, Casner et al. 2014, Casner et al. 2016, Weyer 2007).

Mithin wirft der Einsatz adaptiver Systeme insgesamt eine Reihe schwerwiegender Fragen nach soziotechnischer Systemgestaltung und dem menschengerechten Umgang mit diesen Systemen auf. Als einer dieser Fragen, die bislang kaum Beachtung gefunden hat, der Frage danach, wie die weitgehende Undurchschaubarkeit und Ungewissheit des Verhaltens von Software-Agenten oder MAS sich in der Ko-Aktion mit diesen Systemen auf psychische Belastungen ihrer Nutzer auswirkt, soll im folgenden näher nachgegangen werden.

3 Ein relationales Modell der Stressgenese

Im Zuge der Abkehr von tayloristischen Arbeitsstrukturen mit ihrer stark ausgeprägten horizontalen und vertikalen Arbeitsteilung und Spezialisierung, die sich wegen schwerfälliger Planung und Koordination einer zunehmend dynamischen Umwelt nicht mehr gewachsen zeigen, entstehen komplexe Arbeitsaufgaben, die nur durch selbstorganisierte Kooperation in kleinen Gruppen zu bewältigen sind. Dadurch ergeben sich angereicherte, vielseitige und sinnhafte Tätigkeiten, die hinsichtlich Methoden und Abläufen in vergleichsweise großer Autonomie ausgeübt werden können. Zudem hat der Umfang an projektförmig organisierter Wissensarbeit beträchtlich zugenommen mit weiter wachsender Tendenz. Gerade diese Arbeit erfordert auch selbstorganisierte Kooperation. Nach dem damaligen arbeitswissenschaftlichen Erkenntnisstand entsprachen diese Arbeitsformen weitgehend den Modellvorstellungen guter menschengerechter Arbeit (Ulich 1994). Zugleich zeigen sich aber gerade bei diesen neuen Arbeitstätigkeiten oft auch hohe psychische Belastungen und Stressreaktionen, die auf Unstimmigkeiten in diesen Vorstellungen verwiesen.

Vor diesem Hintergrund entstand das in der Stressforschung weithin beachtete »Job Demand and Control«-Modell der Stressgenese von Karasek & Theorell (1990), das gleichfalls auf den Zusammenhang von Arbeitsanforderungen und Handlungsspielraum fokussiert. Diesem Modell zufolge hängen Stressreaktionen wesentlich von zwei Faktoren ab, hohen herausfordernden Arbeitsanforderungen unter Zeitdruck und dem Ausmaß des Handlungsspielraums in der Arbeit. Autonomie hat dabei eine moderierende Wirkung: Gesteigerte Arbeitsanforderungen können geringere Stressreaktionen auslösen, wenn die Arbeitsperson ihren Arbeitsprozess hinsichtlich der Arbeitsmittel und Vorgehensweise zu kontrollieren vermag. So könnten selbst hohe Arbeitsanforderungen in Verbindung mit weitreichender Kontrolle über den Prozess zu Wohlbefinden und persönlicher Entwicklung führen, soweit sie Lernprozesse ermöglichten.

Da damit die angedeuteten Unstimmigkeiten in diesem Modellansatz der Ausbalancierung von Anforderungen und Kontrolle nicht wirklich überwunden werden, haben Maslach & Leiter (1997) den Ansatz erweitert, indem sie in ihrem Modell berücksichtigen, dass Stressreaktionen von einem ganzen Bündel an Diskrepanzen zwischen Arbeitsanforderungen und verfügbaren Ressourcen abhängen. Diese Diskrepanzen können, wenn sie länger andauern, zu »Burnout« als einem Zustand physischer und emotionaler Erschöpfung und reservierter Gleichgültigkeit führen. Im einzelnen richten sie dabei ihr Augenmerk auf die Diskrepanz zwischen Arbeitsbelastung und Ressourcen, auf mangelnde Kontrolle, auf unzureichende soziale Anerkennung und unfaire Behandlung, auf den Verlust unterstützender sozialer Beziehungen in der Arbeit sowie auf individuelle und organisationale Wertkonflikte. Auf diese Weise haben sie nicht nur die Bedürfnisse und Ressourcen der Arbeitspersonen im Arbeitsprozess im Blick, sondern vor al-

lem auch die Bedingungen der Arbeitsumgebung, in die er eingebettet ist (vgl. Kira 2002, 2003).

Dieser letzte Modellansatz kommt der hier vertretenen Erklärung von Stressgenese schon sehr nahe, weist aber in einer wichtigen Hinsicht noch Defizite auf. Das Kernproblem ist, dass dabei in Betracht gezogene Ressourcen, insbesondere etwa der Grad an Autonomie, *per se*, ohne Rücksicht auf situative Umstände, als Ressourcen aufgefasst werden. Gerade Wissensarbeit unterscheidet sich aber in verschiedener Hinsicht von industrieller Handarbeit: Sie ist oft hoch komplex und erfordert in der Regel vielseitige Fähigkeiten der Problemlösung, etwa zur Erkundung noch unbekannter Lösungen, die gezielt nach wechselnder Kooperation mit anderen Experten oder gar Kunden verlangen. Damit ist Wissensarbeit in hohem Maße auf individuelle Erfahrung und persönliches Können angewiesen. Infolgedessen muss ein adäquater Modellansatz im Kern eine relationale Perspektive auf Arbeitsanforderungen und Ressourcen eröffnen. Mögliche Ressourcen dürfen nicht mehr nur als solche bestimmt und in Betracht gezogen werden, sondern nur in Relation zur jeweiligen Arbeitssituation und den übrigen Bedingungen, unter denen sie aktiviert werden.

Mit der Perspektive »widersprüchlicher Arbeitsanforderungen« (Moldaschl 2005) wurde ein theoretischer Zugang zu diesen Zusammenhängen gefunden und als relationales Modell der Stressgenese operationalisiert. Diesem Modell zufolge werden arbeitsbedingte Belastungen verursacht durch Widersprüche oder Diskrepanzen zwischen in einer Situation gegebenen Arbeitsanforderungen, eingespielten Regeln und verfügbaren Ressourcen. Stressreaktionen entstehen dann, wenn (Wissens-)Arbeiter Widersprüche zwischen Anforderungen, Regeln und Ressourcen zu bewältigen haben, die das Erreichen der Arbeitsziele behindern oder erschweren, dadurch Gesundheit wie Motivation beeinträchtigen. Insbesondere können damit über individuelle Gegebenheiten hinaus auch organisationale Aspekte wissensbasierter Projektarbeit analysiert werden.

In dieser relationalen Sicht werden Ressourcen als wirksame Mittel betrachtet, die (Wissens-)Arbeiter in einer Arbeitssituation zu aktivieren vermögen, um ihre jeweiligen Ziele zu erreichen. Ob etwas als Resource dienen kann oder nicht und welche Wirksamkeit sie dabei entfaltet, hängt dann von den Kontext- und Rahmenbedingungen ab, unter denen die Arbeit ausgeführt wird. Ressourcen lassen sich mithin nur im Gebrauch selbst bestimmen: Einige verfügbare Mittel mögen als Resource genutzt werden können, um Widersprüche in den Anforderungen aufzulösen, sie können aber auch unter anderen Umständen als Stressoren wirken. Beispielsweise kann soziale Unterstützung bei der Lösung einer Aufgabe helfen, unter besonderem Zeitdruck kann eben dies aber auch hinderlich sein. Ebenso mag Autonomie als Ressource bei der Verbesserung von Arbeitsprozessen dienen, sie bleibt aber ohne Wirkung, wenn andere Bedingungen sie nicht zu nutzen erlauben. Mit Blick auf Operationalisierung wird darauf fokussiert, dass psychischer Stress vor allem durch notwendig gewordene Zusatzarbeit angezeigt

wird und dass darüber hinaus in erhöhtem Maße Bewältigungskompetenzen im Umgang mit unausgeglichenen bzw. widersprüchlichen Arbeitsanforderungen, etwa Missverhältnissen von Aufgaben und verfügbaren Ressourcen, hohem Zeitdruck oder häufigen Arbeitsunterbrechungen, in Anspruch genommen werden (Brödner 2009; vgl. Abb. 3).

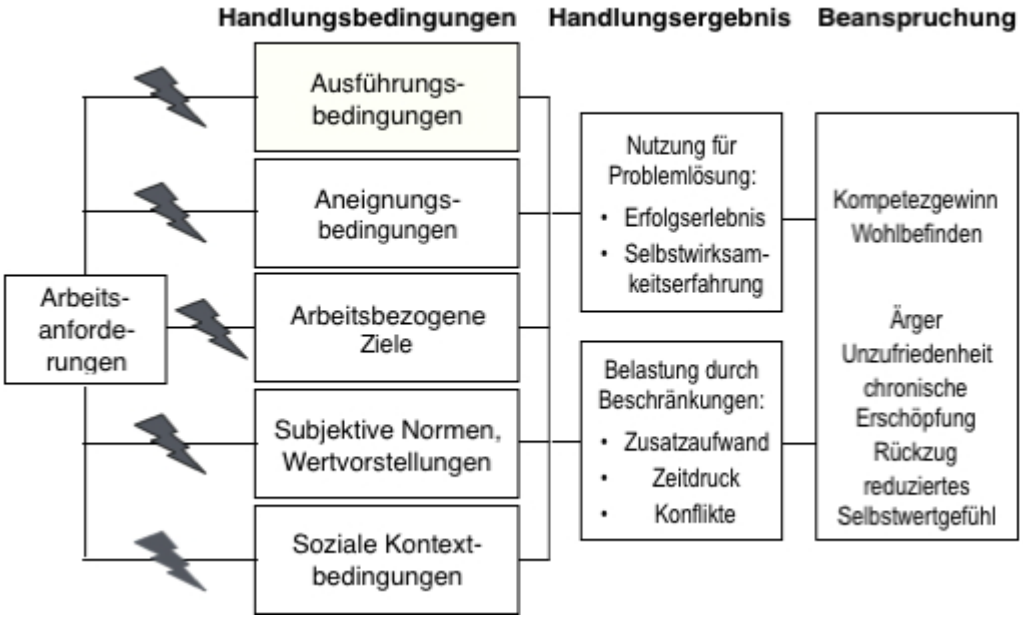


Abb. 3: Ein relationales Modell der Stressgenese (eigene Darstellung, angelehnt an Brödner 2009 und Gerlmaier & Latniak 2011)

Von besonderer Bedeutung ist in diesem Zusammenhang zudem, dass die meisten der dabei betrachteten menschlichen Ressourcen *generativer* Natur sind, d.h. dass sie im Gebrauch nicht wie physische Ressourcen verschleißen, sondern eher wachsen. So gehören zu den generativen Ressourcen etwa persönliche Fähigkeiten und Kenntnisse sowie soziale Beziehungen wie Vertrauen oder Verpflichtung. Dagegen können sie auch leicht zerstört werden, wenn Gelegenheiten und Fähigkeiten zur Erholung von Stress fehlen oder zu stark eingeschränkt sind. Gute soziotechnische Systemgestaltung hat daher dafür zu sorgen, dass sich generative Ressourcen im Arbeitsprozess mindestens in dem Maße neu bilden und entfalten können wie sie anderweitig zerstört werden.

4 Folgerungen für Arbeitsbelastungen und nachhaltige Gestaltung von Ko-Aktion

Das dargestellte relationale Modell der Stressgenese ist in einer Vielzahl von Projekten zur Analyse und Gestaltung von Wissensarbeit, insbesondere auch computerunterstützter kognitiver Arbeit, erfolgreich verwendet, erprobt und in Einzelheiten der Operationaliserung verfeinert worden (Gerlmaier & Latniak 2011).

Schon beim Einsatz von und Umgang mit herkömmlichen Computersystemen häufig angetroffene typische Beispiele für widersprüchliche und folglich Stress generierende Arbeitssituationen sind beispielsweise:

- Widersprüche zwischen Aufgaben und Ausführungsbedingungen oder Lernmöglichkeiten beschränken die Handlungsregulation bzw. die Aneignung nötigen Wissens: z.B. entsteht Zusatzaufwand durch gebrauchsuntaugliche Werkzeuge.
- Widersprüchliche Ziele der Projektarbeit stürzen Arbeitspersonen in Loyalitätskonflikte.
- Widersprüche zwischen aufgabenbezogenen und persönlichen bzw. professionellen Werten verursachen Wertkonflikte.

- In der Ko-Aktion mit adaptiven Systemen (sog. »intelligenten Agenten«) sehen sich (Wissens-)Arbeiter mangels ausreichender Kontrolle über ihr Arbeitsmittel (wie oben ausgeführt) mit weiteren gravierenden widersprüchlichen Arbeitsanforderungen konfrontiert, insbesondere dann, wenn sie infolge indirekter Steuerung und »interessierter Selbstgefährdung« (Peters 2011) unter starkem Leistungsdruck stehen. Als System-Nutzer sind sie dann infolge der Intransparenz des Systemverhaltens und der Unsicherheit seiner Ergebnisse
  - zu blindem Vertrauen in die Systemausgaben verdammt, deren Validität und Qualität sie jedoch nicht zu beurteilen vermögen – ein Widerspruch zwischen Arbeitsanforderung und Ausführungsbedingungen,
  - daran gehindert, sich die Systemfunktionalität für instrumentelles Handeln hinreichend anzueignen (wegen Verletzung des Grundsatzes der Ewartungskonformität) – ein Widerspruch zwischen Arbeitsanforderung und Aneignungsbedingungen,
  - oft auch im Unklaren darüber gelassen, wie mit Fehlleistungen umgegangen und wer dafür verantwortlich gemacht wird – ein Widerspruch zwischen Arbeitsanforderung und Werten,
  - aufgrund der Differenz zwischen erwartetem und tatsächlichem Systemverhalten bei zugleich hoher Leistungserwartung auch hier zu beträchtlichem Zusatzaufwand veranlasst sowie vor Loyalitäts- und Wertkonflikte gestellt.

In Anbetracht der begrenzt erscheinenden Möglichkeiten vollständiger Automatisierung, dazu geeignet, ganze Arbeitsprozesse komplett zu ersetzen, ist auf absehbare Zeit mit der Notwendigkeit des Zusammenwirkens von Wissensarbeitern mit adaptiven Systemen zu rechnen. Dafür können nun, gestützt auf das vorgestellte relationale Modell der Stressgenese, Schlussfolgerungen für die soziotechnische Arbeitsgestaltung der Ko-Aktion gezogen werden. Um dazu nachhaltige, produktive und zugleich sozialverträgliche Arbeitssysteme zu schaffen, in denen Wissensarbeiter durch fortgeschrittene Computersysteme adäquat unterstützt werden, sind eine Reihe von Forderungen zu erfüllen.

Zunächst sollte für Aufgaben, deren Bewältigung weiterhin qualifizierte (Wissens-)Arbeit erfordert, der Einsatz adaptiver Systeme wegen der damit ver-

bundenen widersprüchlichen Arbeitsanforderungen mit ihren hohen Stressrisiken, insbesondere wegen der kaum einschätzbaren Qualitätsprobleme bei den System-Trainingsdaten und wegen der Unsicherheiten ihrer Ergebnisse, überhaupt vermieden werden. Dieser Verzicht empfiehlt sich jedenfalls, solange die Systeme ihr Verhalten und die gelieferten Resultate nicht auf Verlangen im einzelnen zu erklären vermögen. Zwar wird an der Selbsterklärung des Systemverhaltens derzeit geforscht, aber mit bei weitem noch nicht befriedigenden Ergebnissen (vgl. Kindermans et al. 2017; Park et al. 2017). Einstweilen sollte daher der Einsatz adaptiver Systeme, um diesen besonderen Risiken zu entgehen, auf solche Aufgaben und Einsatzfälle beschränkt werden, die sich mit hinreichender Ergebnissicherheit vollständig automatisieren lassen.

Dementsprechend sollten sich weitere Forschungsanstrengungen zu adaptiven Systemen auf die Entwicklung und Implementierung von Einrichtungen konzentrieren, die auf Verlangen fragwürdiges Systemverhalten nachvollziehbar zu erklären vermögen. Dabei ist wichtig, den Fokus und Detaillierungsgrad der Erklärung situationsabhängig wählen zu können derart, dass für die Nutzer Lern- und Aneignungsmöglichkeiten erweitert werden.

Des weiteren sollte, um widersprüchliche Arbeitsanforderungen und daraus resultierende Belastungen durch den Einsatz adaptiver Systeme möglichst einzuschränken, in Fällen des Systemversagens Verantwortung nur aufgrund sorgfältiger Untersuchungen und umfassender Analysen und nicht vorschnell menschlichem Versagen zugeschrieben werden. Es muss legitim sein, unzureichend erklärte Systemresultate zu umgehen. Zudem müssen auch adaptive Systeme haftbar gemacht werden können. Dafür sind ausreichende Untersuchungskapazitäten einzurichten.

Darüber hinaus sollten adaptive Systeme, wie übrigens frühere Hochrisiko-Artefakte auch schon, einer öffentlich kontrollierten Zertifizierung unterliegen. Dabei muss der Beurteilung der Qualität der Erklärungskomponenten hinsichtlich der Lern- und Aneignungsmöglichkeiten durch die Nutzer besondere Aufmerksamkeit zuteil werden.

5 Ausblick: Perspektiven künftiger soziotechnischer Arbeitsgestaltung

Beim gegenwärtigen Entwicklungsstand bleibt einstweilen nur die Wahl, Arbeitsprozesse mittels adaptiver Systeme entweder vollständig zu automatisieren und ihnen trotz aller Risiken des Versagens blindlings zu vertrauen oder aber auf deren Einsatz zu verzichten. Damit stellt sich die Frage nach alternativer Gestaltung von produktiven und sozialverträglichen Arbeitssystemen umso dringlicher.

Im Kern müsste dazu die vorherrschende ideologische Fixierung und trügerische Hoffnung auf ›künstliche Intelligenz‹ und das datengetriebene ›maschinelle Lernen‹ überwunden werden. Anstatt Fragen nach der Beschaffung möglichst umfangreicher Datenbestände und ihrer automatischen Verarbeitung nachzugehen,

deren Brauchbarkeit und Nutzen höchst fragwürdig ist, wäre es weit angemessener, zunächst einmal drängende Probleme zu identifizieren, die Produktions- und Dienstleistungsprozesse tatsächlich in ihrer Leistungsfähigkeit beeinträchtigen. Statt nach Problemen zu suchen, auf die altbekannte Methoden und deren Vergegenständlichung in leistungsstarken Software-Artefakten passen, sollte eher danach gesucht werden, wie zur Lösung von Problemen die besonderen Fähigkeiten und Erfahrungen menschlicher Experten mit der Leistung von Computersystemen wirksam unterstützt werden können. Dementsprechend sollten Forschungsanstrengungen auf die Entwicklung fortgeschrittener Computersysteme ausgerichtet werden, die qualifizierte (Wissens-)Arbeit unterstützen und effektiver machen, aber nicht ersetzen (vgl. auch Norman 2017; Brödner 2019b).

Zur Lösung solcher Probleme können dann organisatorische ebenso wie technische Lösungsansätze mit entsprechender Personalentwicklung in Betracht gezogen werden. Wie bisherige Ansätze zur Leistungssteigerung von Organisationen, insbesondere auch der Einsatz von Computersystemen, lehren, sind organisatorische Veränderungen sogar meist günstiger und effektiver.

6 Literatur

Andelfinger, U. (1997): Diskursive Anforderungsanalyse. Ein Beitrag zum Reduktionsproblem bei Systementwicklungen in der Informatik, Frankfurt/M: Peter Lang

Bainbridge, L. (1983): Ironies of Automation, 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: P. Turner & S. Turner (eds.): European Conference on Cognitive Ergonomics, ECCE '12, Edinburgh (UK), August 28-31, 2012, 65-71

Bradshaw, J.M.; Feltovich, P. & Johnson, M. (2011): Human-Agent Interaction, in: G.A. Boy (ed.): The Handbook of Human-Machine Interaction. A Human-Centered Design Approach, Boca Raton (FL): CRC Press

Bradshaw, J.M. & Hutchinson, F. (eds.) (1997: Software Agents, Cambridge (MA): MIT Press

Brödner, P. (2019a): Coping with Descartes' Error in Information Systems, AI & Society Journal of Knowledge, Culture and Communication 34, 203-213

Brödner, P. (2019b): Grenzen und Widersprüche der Entwicklung und Anwendung ›Autonomer Systeme‹, in: H. Hirsch-Kreinsen & A. Karačić (Hg.): Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt, Bielefeld: transcript, 69-97

Brödner, P. (2017): Die dritte Welle der »automatischen Fabrik« – Mythos und Realität semiotischer Maschinen, in: G. Banse; U. Busch & M. Thomas (Hg.): Digitalisierung und Transformation. Industrie 4.0 und digitalisierte Gesellschaft, Abhandlungen der Leibniz-Sozietät der Wissenschaften Band 49, Berlin: trafo Wissenschaftsverlag 2017, 165-184

Brödner, P. (2009): Sustainability in Knowledge-Based Companies, in: P. Docherty; M. Kira & R.
Shani (eds.) (2009): Creating Sustainable Work Systems. Developing Social Sustainability, London: Routledge, 53-69

Broy, M. (Hg.) (2010): Cyber-physical systems. Innovation durch softwareintensive eingebettete Systeme, Berlin Heidelberg: Springer

Burges, C.J.C. (1998): A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery 2, 121-167

Casner, S.M.; Hutchins, E.L. & Norman, D. (2016): The Challenges of Partially Automated Driving, CACM 59 (5), 70-77.

Casner, S.M.; Geven, R.W.; Recker, M.P. & Schooler, J.W. (2014): The Retention of Manual Flying Skills in the Automated Cockpit, Human Factors 56 (8), 1506-1516.

EU Kommission (2020): Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, Brüssel: EU-Kom

Floyd, C. (2002): Developing and Embedding Autooperational Form, in: Y. Dittrich; C. Floyd & R. Klischewski (Eds.): Social Thinking – Software Practice, Cambridge (MA): MIT Press, 5-28

Fodor, J. 1968: Psychological Explanation. An Introduction to the Philosophy of Psychology, New York: Random House

Foerster, H. von (1993): Wissen und Gewissen, Frankfurt/M: Suhrkamp

Gerlmaier, A. & Latniak, E. (Hg.) (2011): Burnout in der IT-Branche. Ursachen und betriebliche Prävention, Kröningen: Asanger Verlag

Hubig, C. (2019): Haben autonome MaschinenVerantwortung?, in: H. Hirsch-Kreinsen & A. Karačić (Hg.): Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt, 275-298

Jeschke, S. (2015): Auf dem Weg zu einer »neuen KI«: Verteilte intelligente Systeme, Informatik Spektrum 38 (1), S. 4-9

Kabel, D.B.; Riley, J.M.; Tan, K.-W. & Endsley, M.R. (2001): On the Design of Adaptive Automation for Complex Systems, Int. Journal of Cognitive Ergonomics 5 (1), 37-57

Karasek, R. & Theorell, T. (1990): Healthy Work. Stress, Productivity, and the Reconstruction of Working Life, New York: Basic Books

Kindermans, P.-J.; Schütt, K.T.; Alber, M.; Müller, K.-R.; Erhan, D.; Kim, B. & Dähne, S. (2017): Learning how to Explain Neural Networks: PatternNet and PatternAttribution, arXiv:1705.05598v2

Kira, M. (2003): From Good Work to Sustainable Development – Human Resource Consumption and Regeneration in the Post-Bureaucratic Working Life, Stockholm: KTH

Kira, M. (2002): Moving from Consuming to Regenerative Work, in: P. Docherty; M. Kira & R.

Shani (eds.) (2009): Creating Sustainable Work Systems. Emerging Perspectives and Practice, London: Routledge, 29-39

Kriesel, D. (2005): Ein kleiner Überblick über Neuronale Netze, http://www.dkriesel.com/science/neural_networks

Lovelock, J. (2019): Novacene. The Coming Age of Hyperintelligence, London: Allen Lane

Marcus, G. (2020): The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence, https://arxiv.org/pdf/2002.06177.pdf

Maslach, C. & Leiter, M.P., 1997: The Truth about Burnout. How Organizations cause Personal Stress and What to Do about It, San Francisco (CA): Jossey-Bass

McCarthy, J. (1955): A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

McCay, D. (2003): An Example of Inference Task: Clustering, in: D. McCay (ed.): Information Theory, Inference, and Learning Algorithms, Cambridge: Cambridge University Press, 284–292

Mittelstadt, B. D.; Allo, P.; Taddeo, M.; Wachter, S. & Floridi, L. (2016): The Ethics of Algorithms: Mapping the Debate, Big Data & Society 3 (2), 1-21

Moldaschl, M. (ed.) (2005): Immaterielle Ressourcen. Nachhaltigkeit von Unternehmensführung und Arbeit I, München: Hampp

Nake, F. (1992): Informatik und die Maschinisierung von Kopfarbeit, in: W. Coy et al. (Hg.): Sichtweisen der Informatik, Brauchschweig-Wiesbaden: Vieweg, 181-201

Nilsson, N.J. (1998): Introduction to Machine Learning, http://robotics.stanford.edu/~nilsson/ML-BOOK.pdf

Nöth, W., 2002: Semiotic Machines, Cybernetics and Human Knowing 9 (1), 5-22

Norman, D.A. (2017): Design, Business Models, and Human-Technology Teamwork, Research Technology Management 60 (1), 26-30

Norman, D.A. (1994): How Might People Interact with Agents, CACM 37 (7), 68-71 Park, D.H.; Hendricks, L.A.; Akata, Z.; Schiele, B.; Darrell, T. & Rohrbach, M. (2017): Attentive Explanations: Justifying Decisions and Pointing to the Evidence, arXiv:1612.04757v2 Peirce, C.S., 1983: Phänomen und Logik der Zeichen, Frankfurt/M: Suhrkamp

Peters, K. (2011): Indirekte Steuerung und interessierte Selbstgefährdung, in: N. Kratzer; W.

Dunkel; K, Becker & S. Hinrichs (Hg.): Arbeit und Gesundheit im Konflikt, Berlin: edition sigma, 105-122

Putnam, H. 1960: Minds and Machines, in: Hook, S. (ed.): Dimensions of Mind, New York (NY): New York University Press

Russell, S.; Dewey, D. & Tegmark, M. (2015): Research Priorities for Robust and Beneficial Artificial Intelligence, AI Magazine Winter 2015, 105-114

Russell, S. & Norvig, P. (2009): Artificial Intelligence: A Modern Approach, 3rd. ed., Upper Saddle River (NJ): Pearson

Ulich, E. (1994): Arbeitspsychologie, 3. Aufl., Stuttgart: Schäffer Poeschel and Zürich: vdf Weber, M.I.; Mönning, A.; Hummel, M.; Weber, E.; Zika, G.; Helmrich, R.; Maier, T. & Neuber-Pohl, C. (2016): Wirtschaft 4.0 und die Folgen für Arbeitsmarkt und Ökonomie, IAB Forschungsbericht 13/2016

Weyer, Johannes (2007): »Autonomie und Kontrolle. Arbeit in hybriden Systemen am Beispiel der Luftfahrt«, in: Technikfolgenabschätzung – Theorie und Praxis 16 (2), S. 35-42.

Wooldridge, M. 2002: An Introduction to Multi-Agent Systems, Chichester: Wiley

↑Inhalt↑

Das Produktivitätsparadoxon der Computertechnik

1 Einführung: Computer als universelle Rationalisierungsmaschine?

Rationalisierung ist ein schillernder Begriff. Die Vieldeutigkeit steckt schon im lateinischen Wortstamm ›*ratio*‹, der je nach Kontext sowohl ›Berechnung‹ als auch ›Vernunft‹ bedeuten kann. Ersteres trifft auf Computer genau zu: Sie führen nichts weiter als berechenbare Funktionen aus. In diesem Sinne können Computer und vernetzte Computersysteme zurecht als komplexe Maschinen zur Rationalisierung aufgefasst werden. So heißt im Angelsächsischen die wissenschaftliche Lehre, die deren Entstehung und Betrieb sowie ihre logisch-mathematischen und physikalischen Grundlagen zum Gegenstand hat, ebenso zutreffend ›Computer‹ bzw. ›Computing Science‹.

Im üblichen Sprachgebrauch wird unter Rationalisierung jedoch noch etwas anderes verstanden. Häufig wird mit Rationalisierung zunächst in einengender Weise ein Bündel von Maßnahmen zur Kostensenkung und zur Reduktion von lebendiger Arbeit bezeichnet, die in der Folge zu höherer Arbeitsproduktivität und Abbau von Arbeitsplätzen führen. Diese enge Sicht wird aber der tatsächlichen vielseitigeren wirtschaftlichen Dynamik industrieller Entwicklung seit Ende des 18. Jahrhunderts nicht gerecht, dazu muss der Rationalisierungsbegriff weiter gefasst werden. *Rationalisierung* meint dann in diesem weit gefassten Sinn, Arbeits- und Wertschöpfungsprozesse nach jeweils verfügbarem Wissen so zu gestalten, dass sozial und wirtschaftlich gewünschte Effekte mit dem jeweils geringst möglichen Aufwand erzielt werden. Das umfasst nicht nur zweckmäßige Mittel, sondern bezieht auch die »Vernünftigkeit von Zwecken« in die Betrachtung ein (so bereits Weber 1921). Mithin richtet sich Rationalisierung auf in der Gesellschaft als relevant erachtete Ziele, etwa Produktivität und Wettbewerbsfähigkeit von Unternehmen oder auch sozialverträgliche Gestaltung von Arbeit und Technik, und hängt dabei von jeweils vorgefundenen Umständen ab.

So hat uns die durch Zwänge der Kapitalverwertung angetriebene Industrialisierung, anfangs aufgrund ausgeprägter Arbeitsteilung in eng spezialisierte Verrichtungen bei Hand- und Kopfarbeit gleichermaßen (Smith 1776, Babbage 1832), im weiteren Verlauf auch durch massiven Einsatz von energie- und stoffumwandelnden Maschinen (»Kraft- und Arbeitsmaschinen«) eine enorme Steigerung der Arbeitsproduktivität und mit ihr beträchtlichen Wohlstand beschert. Seit 1850 hat

sich die Arbeitsproduktivität in entwickelten Gesellschaften um mindestens den Faktor 20 erhöht, bei gleichzeitiger Verdoppelung der Lebenserwartung, erheblichem Wachstum der Bevölkerung und starker Urbanisierung. Begleitet wurde diese in der Kulturgeschichte singuläre Entwicklung durch die ›Verwissenschaftlichung‹ der (Re-)Produktion, d.h. durch deren analytische Durchdringung und Explikation von Erfahrung in begriffliches Wissen. Als Folge und weitere Triebkraft zugleich wachsen Bildungs- und Forschungseinrichtungen und zeichenbasierte Wissensarbeit breitet sich überall stark aus. Heute ist der Wert der spezifischen Verwendung von Wissen in Produkten oder Diensten oftmals schon größer als der eingesetzte Materialwert; ohne adäquates Wissen gibt es weder Innovation noch produktive Arbeit.

Vor dem Hintergrund dieser Entwicklung wird nun heute die produktivitätssteigernde Wirkung des Einsatzes von energie- und stoffumwandelnden Maschinen zumeist umstandslos auf den Einsatz von Computern als dominanter Basistechnik im Umgang mit zeichenbasierter Wissensarbeit in Wertschöpfungsprozessen übertragen (heute irreführend als ›Digitalisierung‹ bezeichnet). So werden Computersysteme gemeinhin als technische Artefakte betrachtet, mittels derer jedwede Wissensarbeit neu gestaltet und leistungsfähiger, insbesondere produktiver gemacht werden kann. Dank ihrer Programmierbarkeit und vielfältigen Einsatzfähigkeit werden sie auch als universell nutzbare ›enabling technology‹ gesehen, von der hohe Effizienzgewinne erwartet werden.

So werden seit inzwischen über vier Dekaden immer wieder große Summen in den Einsatz von Computersystemen und deren Vernetzung investiert in der Hoffnung auf ebenso große Steigerungen der Arbeitsproduktivität, meist freilich ohne die erwarteten Wirkungen tatsächlich zu überprüfen. Regierungen legen stets aufs neue gut ausgestattete Förderprogramme auf, um die Innovationskraft und Wettbewerbsfähigkeit von Unternehmen zu sichern und die Leistung öffentlicher Dienste zu steigern. Manche Beobachter sprechen in diesem Kontext wiederholt sogar von einem entstehenden »globalen Informationsraum«, der sich durch die Nutzung via Internet global vernetzter Computer bilde und einen »Produktivkraftsprung« bewirke (z.B Boes et al. 2014). Entsprechend wird auch immer wieder technologische Arbeitslosigkeit befürchtet (›Computer als Jobkiller‹), diesmal auch bei qualifizierten Wissensarbeitern. Während in den 1980er Jahren im Zusammenhang mit Forschungen zu wissensbasierter »symbolischer künstlicher Intelligenz (KI)« von der »menschenleeren Fabrik« die Rede war, bestimmen heute ›maschinelles Lernen‹ und ›künstliche neuronale Netze‹, sog. subsymbolische ›KI‹, den öffentlichen Diskurs um Folgen der sog. ›Digitalisierung‹ für Arbeit und Beschäftigung.

Dem stehen freilich empirische Befunde in großer Zahl und erdrückender Aussagekraft entgegen, die auf einzelwirtschaftlicher Ebene nur in einzelnen Fällen unter jeweils bestimmten Kontextbedingungen Produktivitätssteigerungen konstatieren und gesamtwirtschaftlich trotz massiver Investitionen, die seit

längerem diejenigen in Produktionstechnik übersteigen, kaum zusätzliche produktivitätssteigernde Effekte nachzuweisen vermögen. Dieses Phänomen wird als Produktivitätsparadoxon der Computertechnik bezeichnet und wurde erstmals von Robert Solow (1987) konstatiert: »You can see the computer age everywhere but in the productivity statistics«.

Diesem Paradoxon will der vorliegende Beitrag nachgehen, indem zunächst die wichtigsten empirischen Befunde zu den Produktivitätseffekten des Computereinsatzes auf gesamt- wie auf einzelwirtschaftlicher Ebene referiert werden. Danach wird den Gründen für die darin aufscheinende offenkundige Selbsttäuschung der Akteure nachgegangen, die im wesentlichen in einem falschen Verständnis der Besonderheiten der Funktionsweise von Computern und ihres Einsatzes zu suchen sind und das Phänomen hinreichend zu erklären vermögen. Abschließend werden Schlüsse gezogen, wie künftig diese Technik besser gestaltet und genutzt werden kann.

2 Das Produktivitätsparadoxon: Empirische Befunde

Seit der berühmt gewordenen Aussage von Solow über die paradoxe, besser gesagt: kontraintuitive Wirkung der Computertechnik auf die Arbeitsproduktivität ist eine Fülle empirischer Untersuchungen über das Phänomen erschienen, zunächst v.a. in den USA (zur Übersicht vgl. etwa Accenture 2015). Aufgrund der vergleichsweise vielfältigeren Analysen und besseren Datenlage wird hier vorwiegend auf US-bezogene Studien zurückgegriffen (zumal die Entwicklung und Anwendung der Computertechnik im jüngeren Verlauf amerikanisch dominiert ist); die Befunde können grundsätzlich aber auch für deutsche und europäische Verhältnisse Geltung beanspruchen. Wie die empirischen Befunde zeigen, ist es dabei aufgrund von Unterschieden in der Auswirkung angemessen, zwischen der gesamt- und der einzelwirtschaftlichen Perspektive zu unterscheiden.

2.1 Gesamtwirtschaftliche Perspektive

Diffusion ab den frühen 1970er Jahren beträchtlich an Fahrt auf und führt im weiteren Verlauf zu massiv wachsenden Investitionen in einer Vielzahl von Sektoren von Produktion und Dienstleistung. In der gleichen Zeit hat sich die Leistung der Hardware-Komponenten (Prozessoren, Speicher, Datentransfer-Kapazität) entsprechend dem Mooreschen ›Gesetz‹ exponentiell gesteigert, was das Preis-Leistungsverhältnis stark verringert und damit die Investitionsdynamik deutlich befeuert. So wird die Hardware bezogen auf ihre Leistung nun ständig billiger – ein Umstand, dem bei der Berechnung der Investitionssummen durch statistische Ermittlung eines Deflators auch Rechnung getragen wird. Andererseits steigt mit der Vielfalt der Einsatzfälle und Aufgaben der notwendige Aufwand für die Entwicklung passender und komplexer werdender Software stark an und macht den Hauptanteil an der Investition aus. Das entgegen allen Erwartungen

Überraschende ist nun, dass exakt in dem Moment zu Beginn der 1970er Jahre, ab dem die Investitionsdynamik in Computertechnik massiv von unter 10 auf über 200 Mrd. $ p.a. (1990) zunimmt, die Arbeitsproduktivität in den Dienstleistungssektoren zu stagnieren beginnt, während sie in der Produktion in gewohntem Maße weiter wächst, was insgesamt das gesamtwirtschaftliche Produktivitätswachstum wegen des zunehmenden Gewichts der Dienstleistungen stark bremst (Brynjolfsson 1993).

Mitte der 1990er Jahre setzt dann ebenso überraschend eine sprunghafte Zunahme der Arbeitsproduktivität von rd. 1% p.a. auf 2,3 % p.a. ein (Stiroh 2002; in Europa deutlich schwächer), was viele Beobachter zu der Vermutung veranlasst, dass sich endlich, nach jahrzehntelanger Verzögerung, die produktivitätssteigernden Wirkungen der Computertechnik auch in der Wirtschaftsstatistik niederschlagen. Das hat sich aber in der Folge als großer Irrtum herausgestellt (Acemoglu et al. 2014). Wie das McKinsey Global Institute (2002) aufgrund von Sektordaten des Bureau of Economic Analysis (BEA) nachweisen kann, lässt sich nämlich der ganze Produktivitätssprung auf einmalige strukturelle Veränderungen in nur 6 Wirtschaftszweigen, darunter ganz überwiegend im Groß- und Einzelhandel (sog. ›Big Box-Strategie‹), zurückführen, die nur sehr bedingt überhaupt mit Computertechnik verbunden sind. Dementsprechend ist die insgesamt nur eine Dekade andauernde Produktivitätszunahme auf über 2% p.a. im Jahre 2005 schon wieder zu Ende. Seitdem liegt sie zunächst bei 1% und ist zuletzt trotz weiterer Investitionen in Computertechnik auf anhaltend hohem Niveau sogar deutlich unter diesen Wert gefallen (Gordon 2014, 2016). Das gilt trotz aller Unterschiede im Detail grundsätzlich auch für andere entwickelte Länder, etwa die G7-Länder, wie nachstehende *Abb. 1* ausweist.

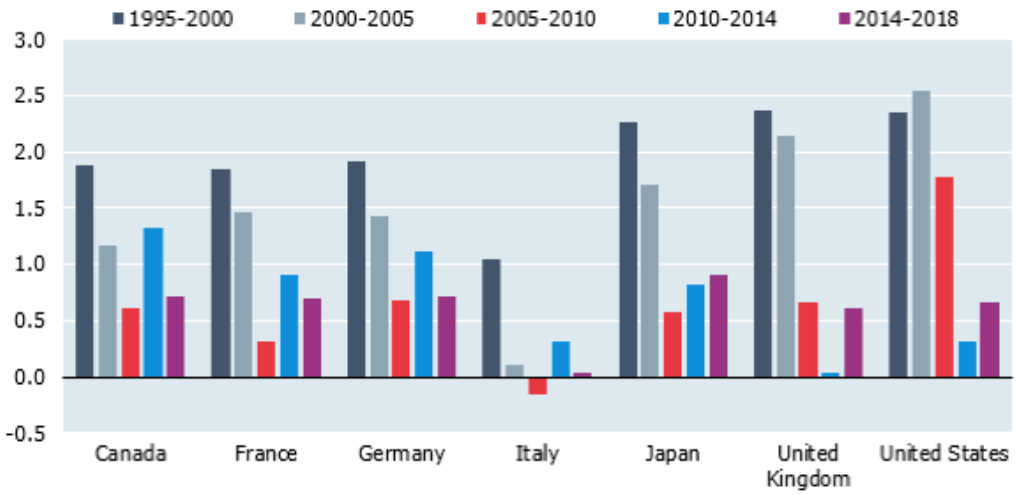


Abb.1: *Entwicklung der Arbeitsproduktivität in G7-Ländern 1995-2018 (OECD 2019)*

Zum besseren Verständnis der Verbreitungsdynamik der Computertechnik ist darauf zu verweisen, dass 1966 mit dem Erscheinen der Mainframe-Rechner IBM 360 und PDP 10 von Digital Equipment, deren ›Virtual Machine‹ genannte Betriebssysteme über Terminals den unmittelbaren Zugang zu Datenverarbeitungskapazität am Arbeitsplatz ermöglichen, ganz neue Arbeitsabläufe bei deren Nutzung eröffnet werden. Außerdem kommen zu Beginn der 1970er Jahre die ersten von Texas Instruments entwickelten Mikroprozessoren auf den Markt und erleben mit dem Z 80 von Zilog 1976 einen Durchbruch. Durch sie wird auch die Entwicklung von Personal Computern möglich (Apple 1976, IBM 1981 mit Betriebssystem von Microsoft, Macintosh von Apple 1984). Grafische Benutzungsoberflächen zur Interaktion mit Computern und gebrauchstaugliche Software für Tabellenkalkulation, Grafik- und Textedition machen Computer nun zu weithin verwendbaren Werkzeugen für Wissensarbeiter. Durch beide Neuerungen wird der Einsatz von Computern aus den Fesseln der von Hohepriestern der Datenverarbeitung bewachten Heiligtümer der Rechenzentren befreit und dem allgemeinen Gebrauch in den vielfältigen zeichenbasierten Arbeitsprozessen von Produktion und Dienstleistung zugänglich. Zudem wird Ende 1982 mit dem Protokoll TCP/IP ein internationaler Quasi-Standard zum Datentransfer zwischen Computern gesetzt, der nach und nach im Zuge des Ausbaus von Übertragungskapazitäten der Telekommunikation die weltweite Vernetzung von Computersystemen mittels kostengünstiger ›Datenfernübertragung‹ ermöglicht.

Bei der Erklärung dieses paradox oder kontraintuitiv erscheinenden Phänomens sinkender Produktivitätszuwächse bei anhaltend sehr hohen Investitionen in Computertechnik als basaler ›enabling technology‹ ist zunächst zu berücksichtigen, dass die gesamtwirtschaftliche Arbeitsproduktivität von einer Vielzahl auch gegenläufiger, sich teilweise wechselseitig kompensierender Einflüsse abhängt. Daher ist es ohne zusätzliche Analysen kaum möglich, einzelne Faktoren wie den Einsatz von Computertechnik soweit zu isolieren, dass deren spezielle Wirkung aufgezeigt werden kann. So zeigt sich in einer Vielzahl von Studien immer wieder, dass der Zusammenhang zwischen Aufwendungen für Computersysteme und deren Wirkungen auf die Arbeitsproduktivität außerordentlich diffus ist (Übersichten über empirische Untersuchungen zur Makro- und Mikroebene geben Potthoff 1998, Accenture 2015). Dessen eingedenk haben sich in jüngerer Zeit die Untersuchungen von gesamtwirtschaftlich orientierten zu einzelwirtschaftlichen Analysen verlagert. Im Umgang einzelner Organisationen mit Computertechnik zeigen sich denn auch erstaunliche Unterschiede.

2.2 Einzelwirtschaftliche Perspektive

Zunächst liefern diesbezüglich verfügbare Daten aus 400 großen US-Unternehmen ein eindrückliches Bild davon, wie höchst unterschiedlich diese Unternehmen produktiven Nutzen aus dem Einsatz von Computersystemen (›IT‹) ziehen (vgl. Abb. 2). Obwohl sie mehr oder weniger alle gleiche oder zumindest funk-

tionsähnliche Systeme nutzen, allerdings in sehr unterschiedlichem Umfang (›IT stock‹), variiert die Arbeitsproduktivität enorm: Z.B. ist sie bei einem Unternehmen am oberen Rand der Punktwolke um den Faktor 4 größer als bei einem am unteren Rand (die Achsen sind logarithmisch skaliert). Zudem ist die Korrelation zwischen dem kumulierten Computereinsatz (›IT stock‹) und der Produktivität insgesamt vergleichsweise schwach. Zwar ist die mit Blick auf die Verfügbarkeit entsprechender Daten getroffene Auswahl der Unternehmen nicht repräsentativ, wegen des fortgeschrittenen Computereinsatzes aber umso instruktiver (Brynjolfsson & Hitt 2000, Brynjolfsson 2003).

Deutlich verweist dieses Bild darauf, dass neben dem Umfang des Einsatzes von Computertechnik noch andere Faktoren die Wirkung auf die Arbeitsproduktivität beeinflussen, etwa die soziale und organisatorische Einbettung der Computersysteme, Managementpraktiken im Umgang mit ihnen oder die Befähigung der Nutzer für deren wirksamen Gebrauch. Weitere Einflüsse sind in einem unzulänglichen Verständnis der Funktionsweise dieser Systeme und ihrer Nutzung, in unpassend definierten Computerfunktionen oder in deren wirksamer Nutzung entgegenstehenden Organisationsformen zu suchen.

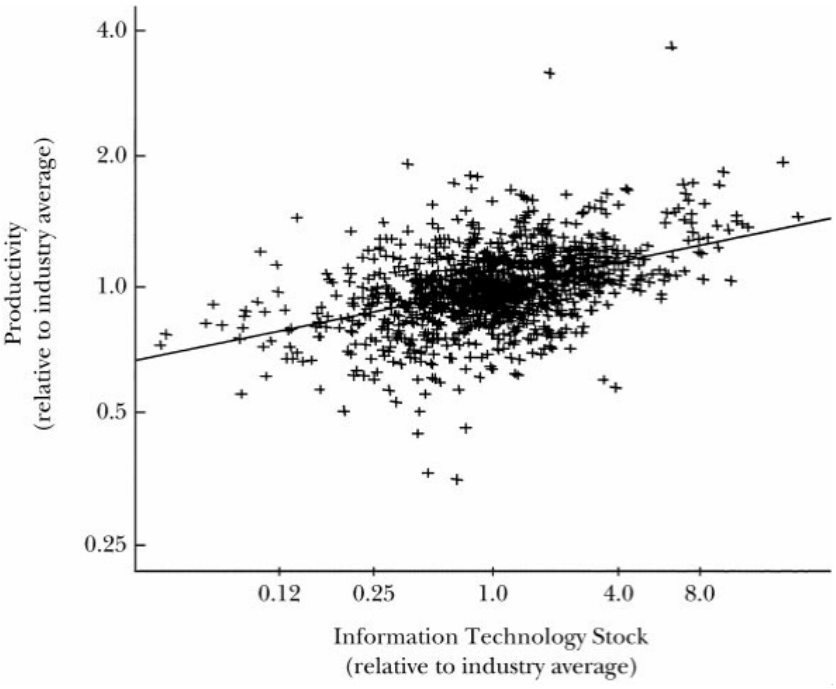


Abb. 2: *Korrelation zwischen Arbeitsproduktivität und Computereinsatz (Daten aus 400 großen US-Unternehmen; Brynjolfsson & Hitt 2000)*

Bereits früh wurde denn auch versucht, diesen Zusammenhängen mittels ökonometrischer Analysen auf die Spur zu kommen. So zeigt sich etwa, dass die Wirkungen isolierter Einsätze von Computersystemen auf die Arbeitsproduktivität für sich genommen eher gering sind. Erst wenn sie mit komplementären organi-

satorischen Veränderungen und Qualifizierungsmaßnahmen zur Ertüchtigung der Benutzer für die neuen Arbeitsstrukturen einhergehen, stellen sich deutliche Steigerungen der Arbeitsproduktivität ein. Aus diesen Analysen ergibt sich insbesondere, dass mögliche Leistungssteigerungen in verschiedener Hinsicht von solchen komplementären Maßnahmen abhängen (Bresnahan et al. 2002, Brynjolfsson 2003):

- Computersysteme vermögen v.a. dann die Leistungsfähigkeit von Unternehmen zu steigern, wenn deren Einführung mit durchgreifender Dezentralisierung, objektorientierter Reorganisation in multifunktionalen Teams und Investitionen in ›Humankapital‹ einhergeht.
- Unternehmen mit derart dezentralisierten Organisationsformen erreichen höhere Produktivität in der Nutzung von Computersystemen als solche, die nur Computer einsetzen.
- So genannte ›intangible assets‹, also z.B. Managementfähigkeiten, Erfahrung oder kollektive Handlungskompetenz, beeinflussen stark den realisierten Nutzen von Computersystemen.
- In der Regel ist der Aufwand für organisatorische Erneuerung und Qualifizierung um ein Vielfaches höher ist als die Ausgaben für Hard- und Software.

Ähnliche Befunde haben sich auch aus eigener Forschung am Institut Arbeit und Technik ergeben. Als in dieser Hinsicht sehr aufschlussreich haben sich Fallstudien zum Einsatz von Produktionsplanungs- und -steuerungssystemen (PPS- bzw. ERP-Systemen) in produzierenden Unternehmen erwiesen. Als integrierte betriebswirtschaftliche Systeme für Materialwirtschaft, Auftragssteuerung, Finanzbuchhaltung und Personalplanung durchdringen sie Organisationen nahezu vollständig und strukturieren deren Arbeitsabläufe. An ihnen lassen sich daher vor allem organisationsbezogene Einsatzprobleme und Herausforderungen wirksamer Aneignung besonders gut studieren.

Beim Einsatz dieser Systeme zeigt sich, dass sieben von zehn Unternehmen einem auf zentrale Planung und Kontrolle ausgerichteten, dabei auf rein technische Funktionalität fixierten Blick auf den Computereinsatz folgen – mit außerordentlich schädlichen Folgen für deren wirtschaftliche Leistung: Projekte der Implementierung überziehen die Zeit- und Kostenbudgets beträchtlich und trotz hoher Kosten verbessern sich wettbewerbsrelevante Leistungsgrößen wie Produktivität, Durchlaufzeiten oder Bestände kaum (was meist ohne Konsequenzen hingenommen wird). Implementierungsprozesse konzentrieren sich ganz überwiegend auf die einmalige Bestimmung der Systemanforderungen zu Beginn, während Fragen der Nutzung und kollektiven Aneignung der Systeme kaum in Betracht gezogen werden. Dabei bleiben viele Funktionen ungenutzt, notwendiges Wissen über Geschäftsprozesse wie über Bedingungen und Folgen kollektiven, systemgestützten Handelns im Organisationszusammenhang ist unter den verschiedenen Akteuren unzureichend vermittelt und oft entstehen fehlerhafte, zu Fehlentscheidungen verleitende Datenhalden (Maucher 1998, 2001).

Nur eine kleine Minderheit von Unternehmen setzt demgegenüber adäquat an organisatorischen Veränderungen an. Mit großem wirtschaftlichem Erfolg richten sie Wertschöpfungsprozesse konsequent am Kundennutzen aus und beginnen mit deren objektorientierter Restrukturierung, um eine dementsprechend angepasste Konzeption und Nutzung von Computersystemen als Arbeitsmittel und Medium der Kooperation zu entwickeln. Zur Bewältigung der organisatorischen Veränderungen und wirksamen Aneignung der Systeme beteiligen sie betroffene Fachleute und organisieren schon in der Konzeptions- und Einführungsphase kollektive Lernprozesse. Nur wenn die Arbeitsweise, mithin die Regeln gemeinsamen Handelns, im Gebrauch der Systeme angemessen verändert werden, lassen sich deren Nutzenpotenziale wirksam aktivieren (Maucher 1998, 2001).

Entsprechend lässt sich der derzeitige Forschungsstand aufgrund einer Metastudie zur Auswertung von über 50 Einzelanalysen wiederum aus den USA wie folgt zusammenfassen:

> *»... the wide range of performance of IT investments among different organizations can be explained by complementary investments in organizational capital such as decentralized decision-making systems, job training, and business process redesign. IT is not simply a tool for automating existing processes, but is more importantly an enabler of organizational changes that can lead to additional productivity gains«* (Dedrick et al. 2003).

Zu einem ganz ähnlichen Schluss gelangen auch Jorgenson et al. (2008) in einer neueren Analyse der Entwicklung der US-Arbeitsproduktivität:

> *»… a wealth of microeconomic evidence emphasizes the complexity of the link from technology to productivity. To leverage information technology investments successfully, firms must typically make large complementary investments and innovations in areas such as business organization, workplace practices, human capital, and intangible capital.«*

2.3 Permanenz der Software-Krise

Die Kosten für Computer Hardware haben seit langem eher sinkende Tendenz, denn deren Leistungsfähigkeit wächst bis in die jüngste Zeit exponentiell; so verdoppelt sich beispielsweise die Rechenleistung von Prozessoren etwa alle zwei Jahre (›Moores Gesetz‹, allerdings nähern sich neuerdings die stofflichen Strukturen physikalischen Grenzen). Die Aufwendungen für Entwicklung, Erprobung und Wartung der Software machen daher schon längst den ganz überwiegenden Teil der Gesamtkosten von Computertechnik aus, v.a. auch, weil Softwaresysteme dank der hohen Hardware-Leistung immer funktionsreicher und komplexer werden können. In Form von immer mehr aufeinander aufbauenden Schichten türmen sich immer vielfältiger zusammengesetzte, abstraktere Funktionen, die jeweils zur Ausführung spezifischer Aufgaben kaskadenförmig auf generische Funktionen tieferer Schichten zurückgreifen. Das gilt insbesondere auch für die komplexen, für den Einsatz in Organisationen bestimmten Computersysteme, die hier im Fokus stehen.

Damit verlagern sich auch die Kostenanteile für die Software selbst mehr und mehr auf deren Wartung und Pflege (einschließlich immer aufwendigerer Fehlerbehebung und Tests, die gleichwohl keine Korrektheit garantieren) sowie die Anpassung an erneuerte Betriebssysteme und veränderte Entwicklungsumgebungen. Wachsende Anteile beansprucht auch die Abwehr und Bewältigung äußerer Angriffe infolge weitreichender Vernetzung. Mit Blick auf die große Bedeutung von Softwarefehlern schätzt die Zeitschrift für Computertechnik IX beispielsweise für das Jahr 2005 den jährlichen Aufwand für Fehlerbeseitigung in deutschen Unternehmen auf 14,4 Mrd. € (36% des IT-Budgets) und die durch Softwarefehler verursachten Verluste auf rd. 84 Mrd. € im Jahr (IX-Studie 1/2006: Software-Testmanagement, Hamburg: Heise). Ähnlich schätzt das US Institute of Standards and Technology (NIST) die Folgekosten von Softwarefehlern im Jahre 2002 auf rd. 60 Mrd. USD (oder 0,6% des US-BIP). Auch wenn das nur Schätzungen sind, geht es hier nicht um ›Peanuts‹.

Bereits im Jahr 1968 wurde auf einer von der NATO gesponserten Konferenz erstmals eine ›Software-Krise‹ ausgerufen (Naur & Randell 1968). Seither haben sich, um der immer wieder neuen Schwierigkeiten im Umgang mit Funktionsvielfalt und Komplexität Herr zu werden, die Methoden des Software Engineering rasant entwickelt, etwa mit systematischen Vorgehensweisen zu weitreichender Abstraktion und Modularisierung, mit abstrakten Datentypen und strukturierter, objektorientierter Analyse, Entwurf und Programmierung oder mit vielfältigen Bemühungen um Weiter- und Wiederverwendung von erprobten Software-Modulen in Gestalt von Programmbibliotheken und Software-Frameworks.

Einerseits haben diese zweifellos hilfreichen methodischen Verbesserungen jedoch im Software Engineering beträchtliche Rebound-Effekte zur Folge: Je zuverlässiger und wirksamer sich diese Methoden zeigen, desto größere und komplexere Projekte werden anzugehen gewagt. Andererseits ändert das an dem in der Softwaretechnik leider gewohnten, verglichen mit anderen Technikfeldern extrem hohen Anteil an Projekten wenig, die entweder vollständig oder zumindest teilweise scheitern, indem sie ihre Zeit- und Geldbudgets bei weitem überziehen, ohne den geplanten Funktionsumfang je zu erreichen (dazu gibt es eine Fülle von Einzelbeispielen, vgl. Landauer 1995; Emam & Koru 2008). Auf Basis der Daten aus über 50.000 Software-Projekten weltweit ergibt sich, dass die Anteile komplett gescheiterter Projekte (um 20% aller Projekte), der teilweise gescheiterten Projekte (um 53%) und schließlich der erfolgreichen Projekte (um 27%) im Zeitverlauf zwar leicht schwanken, sich aber nicht grundsätzlich bessern; sie hängen hauptsächlich von der Projektgröße ab: je größer, umso gefährdeter (Standish Group International 2016). Mehr oder weniger ist in großen Teilen der Software-Entwicklung inzwischen ein Zustand ›rasenden Stillstands‹ erreicht, in dem viel Aktivität ohne substantiellen Fortschritt entfaltet wird.

Diese bei der Nutzung von Computertechnik in Organisationen in besonderem Maße hervortretenden Probleme und trotz aller methodischen Verbesserungen

nun schon seit fünf Dekaden andauernden Krisenerscheinungen im Software Engineering verweisen offenbar auf tiefer liegende systematische Ursachen, die nicht einfach durch einzelne Fehlleistungen zu erklären sind. Nach den vorgestellten empirischen Befunden ist vielmehr zu vermuten, dass wichtige Grundlagen der Fachdisziplin nicht hinreichend geklärt sind, damit grundsätzlich falsche, zu Illusionen führende Annahmen über die Funktionsweise und Nutzungsmöglichkeiten der Systeme wirksam sind, darüber hinaus auch unangemessene Vorgehensweisen beim Projektmanagement zur Gestaltung und Implementation der Systeme zum permanenten Elend beitragen. Diesen Überlegungen wird im folgenden näher nachgegangen.

3 Hintergrund: Einblicke in den Maschinenraum der Computertechnik

3.1 Die Maschine als Fetisch – das verkehrte Verhältnis von Mensch und Maschine

Wenn in entwickelten Gesellschaften über Technik gesprochen oder berichtet wird, dann wird darunter in der Regel schlicht eine Ansammlung technischer Artefakte oder Maschinen verstanden. Sie sind plötzlich wie aus heiterem Himmel da, führen anscheinend ein Eigenleben und ›verändern‹ unser Leben grundlegend, indem sie, wie es heißt, ›etwas mit uns machen‹. Dieses Phänomen wird gelegentlich als »technologischer« oder »Technikdeterminismus« (Lutz 1987) apostrophiert und ist in diesen Tagen, insbesondere im Zusammenhang mit Computertechnik, nach einer Phase zumindest partieller Aufklärung, wieder vorherrschend. Ständig lesen und hören wir von der »Macht der Algorithmen«, die etwas ›prognostizieren‹ (Rückfälligkeit von Straftätern), ›entscheiden‹ (Kreditvergabe) oder ›empfehlen‹ (Routenplanung, Kauf bestimmter Artikel), dass angeblich ›autonome Agenten‹ (z.B. ›selbstfahrende Autos‹) unsere Arbeit übernehmen und uns im Grunde überflüssig machen. Darin kommt ein zwar weit verbreitetes, aber gänzlich fehlgeleitetes Verständnis von Technik zum Ausdruck, das Grundfragen nach dem unterliegenden Menschenbild, unserem Verhältnis zur Welt und zu uns selbst aufwirft.

Im ursprüngliche Wortsinn wird mit *téchne* eine List bezeichnet, die List, Wirkungen der äußeren Natur für eigene Zwecke zu nutzen. Aristoteles sieht darin einen eigenen Teil unserer praktischen Vernunft, die Fähigkeit, etwas Nützliches herstellen zu können, beruhend auf Erfahrung, Übung und Einsicht in Naturverhältnisse. Damit gerät das Verhältnis des Menschen zur äußeren Natur wie auch sein Verhältnis zu sich selbst als sozialem Gattungswesen in den Fokus. In dieser Perspektive erscheint die kulturelle Evolution als die Fortsetzung der natürlichen mit anderen Mitteln: Den Menschen als Produkt natürlicher Evolution kennzeichnet die durch seine bewusste Tätigkeit geschaffene Welt der Kultur, deren

Fortentwicklung sich, getrieben von jeweils gesellschaftlich dominanten Interessen, niederschlägt in Formen der Herstellung und des Gebrauchs von

- Werkzeugen als Mitteln zweckmäßiger Nutzung von Naturkräften (*homo faber*: Werkzeuge sowie energie- und stoffumwandelnde Maschinen vermitteln kausale Wirkungen),
- Zeichen als Mitteln sozialer Interaktion, Kommunikation und Reflexion (›semiotisches Tier‹: Sprache, Schrift und Buchstabendruck vermitteln intentionale Bedeutungen).

So zeigt eine genauere Analyse, dass das auf Artefakte fixierte Technikverständnis in die Irre führt: Technische Artefakte fallen nicht vom Himmel, sondern müssen für bestimmte Zwecke mittels Einsicht in Prozesse der Natur oder sozialer Praxis mühsam konzeptionell entworfen und stofflich hergestellt werden. Als solche sind sie aber bloß tote, nutzlose Gegenstände, solange sie nicht für bestimmte Aufgaben zweckgemäß eingesetzt, mithin dafür angeeignet und praktisch wirksam verwendet werden. Das alles geschieht stets im Spannungsfeld des technisch Machbaren, der Gestaltbarkeit von Natur bzw. sozialer Praxis, und des sozial Wünschenswerten, abhängig von jeweils herrschenden Interessen. In dem auf Artefakte fixierten Technikverständnis verkehrt sich stattdessen das Verhältnis von Menschen zu einander und zur äußeren Natur: Durch Verdinglichung personaler Verhältnisse und Personifizierung von Dingen werden Maschinen zum Fetisch erhoben. Hierin wurzelt die »Macht der Machwerke über die Machenden« (Haug 2005) oder, wie Anders (1973) sagt, die »prometheische Scham« des Menschen vor der Perfektion seiner eigenen Artefakte. Dieser kollektive Wahn führt letztlich zu Infantilisierung und Selbstentmündigung, wie sie etwa in Formen der Verhaltenssteuerung durch Smartphone-Gebrauch bereits zum Ausdruck kommen.

Im Unterschied dazu wird nach allgemeinem professionellen Verständnis Technik definiert als die Gesamtheit von Maßnahmen zur Herstellung und zum Gebrauch künstlicher Mittel für gesellschaftliche Zwecke. Ihr werden damit nicht nur die Artefakte und Sachsysteme selbst zugerechnet, sondern auch deren sozial konstruierte und kulturell vermittelte Herstellung und Anwendung (Ropohl 1991, VDI 1991). Als geronnene Erfahrung verkörpern sie begriffliches, explizites Wissen über die Natur bzw. soziale Praxis und als Arbeitsmittel stellen sie Handlungsanforderungen an ihren Gebrauch, durch den die Artefakte erst ihren Sinn erhalten und in ihrer Qualität zu beurteilen sind. Gerade in den Prozessen der Entwicklung und Herstellung technischer Artefakte sowie ihrer Aneignung zu praktisch wirksamer Verwendung liegen die eigentlichen Probleme von Technik als der ›Anstrengung, Anstrengungen zu ersparen‹, dem Sinn technischen Handelns; eben hierin liegen auch die Wurzeln missbräuchlichen Umgangs.

Vor diesem Hintergrund müssen nun die fundamentalen Unterschiede zwischen klassischen Maschinen der *Energie und Stoffumwandlung* (»Kraft- und Arbeitsmaschinen«, chemische oder biologische Prozesse) und Computern als

semiotischen Maschinen betrachtet werden. Erstere nutzen durchweg Kenntnisse der Thermo- bzw. Elektrodynamik und Mechanik, um zweckmäßig gestaltete Funktionen der *Kraftübertragung* zu realisieren und so Naturkräfte und -effekte nutzen zu können. Im Unterschied dazu greifen Computer in Zeichenprozesse sozialer Praxis ein, sind damit programmgesteuerte, ›Daten‹ verarbeitende Maschinen. Energie- und stoffumwandelnde Maschinen übertragen Kräfte, semiotische Maschinen manipulieren bedeutungslose Zeichenträger.

Die fundamentalen Unterschiede zwischen beiden Maschinenklassen liegen mithin in deren Wirkbereichen, Funktionsweisen und Zwecken. Der *Wirkbereich* von Kraft- und Arbeitsmaschinen (wie auch von artifiziellen chemischen und biologischen Prozessen) liegt in der Natur und nutzt natürliche Kräfte zweckgemäß für Prozesse der Energie- und Stoffumwandlung, während der Wirkbereich semiotischer Maschinen ganz in zeichenbasierter sozialer Interaktion liegt und auf wohl determinierten Funktionen der Verarbeitung von Daten als Zeichenträgern im Rahmen zugrunde liegender Zeichenprozesse beruht. Mit semiotischen Maschinen wird der soziale Raum der Zeichenprozesse und Interaktion nirgends verlassen. Im Unterschied dazu beruht die *Funktionsweise* von Maschinen und Prozessen der Energie- und Stoffumwandlung auf der Einsicht in natürliche Effekte als Ergebnis von Naturerkenntnis und ihr Zweck ist das Nutzen von Naturkräften. Die Funktionsweise semiotischer Maschinen beruht dagegen auf Vorschriften zur Manipulation von Daten, gewonnen durch Analyse, Modellierung und Formalisierung von Zeichenprozessen kognitiver Arbeit und sie dient der Organisation und Koordination kollektiven Handelns. Den zwecks Maschinisierung körperlicher Arbeit geschaffenen mechanischen Funktionen der Kraftübertragung als dem Kern maschineller Energie- und Stoffumwandlung entsprechen dann bei Computern als semiotischen Maschinen die algorithmischen Funktionen für Datenverarbeitung, -speicherung und -transfer zwecks Maschinisierung kognitiver Wissensarbeit.

Beiden Klassen technischer Artefakte gemeinsam ist zunächst, dass beide stets das zur Problemlösung erforderliche explizite methodische und funktionale Wissen vergegenständlichen, das sich der natürlichen analytischen Intelligenz ihrer Konstrukteure verdankt. Beiden Klassen gemeinsam ist ferner ihre enge Verwandtschaft zur Sprache, indem sie auf der Basis von Begriffsbildung und explizitem Wissen absichtsvoll gestaltete, wohl bestimmte Funktionen verkörpern, die durch Menschen in deren Handlungskontext zu interpretieren sind, um sie wirkungsvoll zu gebrauchen (die funktionale ›Sprache‹ der Artefakte). Dabei sind die Wirkungen kraft der wohlbestimmten Funktionen durch die Eingaben determiniert. Um sinnvolle Eingaben machen und deren kausale Wirkungen interpretieren zu können, müssen Handlungen ihres Gebrauchs in der funktionalen ›Sprache der Artefakte‹ zum Ausdruck gebracht werden. Das gilt für alle technischen Artefakte, vom Faustkeil bis zum Computer. Auf der Grundlage dieser Unterschiede und Gemeinsamkeiten können nun die Besonderheiten der Funktions-

weise von Computern als semiotischen Maschinen in Betracht gezogen und die Gründe für ihre verbreitet Mystifizierung aufgezeigt werden.

3.2 Wider die Mystifizierung des Computers

In der Computertechnik führen falsche Bezeichnungen und Metaphern vom Beginn an in die Irre. Seit ihren ersten Implementierungen am Ende des zweiten Weltkriegs bis in unsere Tage mit der Realisierung vermeintlich ›künstlicher Intelligenz‹ und ›maschinellen Lernens‹ mystifizieren sie die semiotische Maschine. Unzählige Berichte sprechen damals vom Computer als »Elektronengehirn«, wie einige ausgewählte Beispiele illustrieren: »30-Tonnen-Elektronengehirn an der Philadelphia Universität denkt schneller als Einstein« *(*Philadelphia Evening Bulletin 15.02.1946), »Giant Brains or Machines That Think« (Buchtitel von E.C. Berkeley 1949), »Can Man Build a Superman?« (Titel des Time Magazine mit Bild des Mark III vom 23.01.1950) oder »The Computer and the Brain« (Buchtitel von J.v. Neumann 1958).

Die Folgen dieser falschen Metaphorik, wie sie auch im äußerst wirkmächtigen sog. ›computational model of the mind‹ zum Ausdruck kommt, sind bis heute schwerwiegend: Damals konzipiert auch Turing (1950) mit der Konstruktion des »imitation game« seinen Turing-Test, der bis heute als das Maß aller Dinge gilt (wenn er auch zur Beurteilung der Frage »can machines think?« gänzlich wertlos ist). Gefangen im damaligen Zeitgeist des Behaviorismus wird damit vom äußerlich beobachtbaren Verhalten einer Maschine auf deren ›Intelligenz‹ geschlossen, wenn Menschen deren Verhalten nicht mehr von dem eines Menschen zu unterscheiden vermögen. Abgesehen davon, dass es sich hier um eine reine Zuschreibung handelt, ist menschliches Urteilsvermögen auch sehr leicht zu täuschen, wie etwa Weizenbaums (1966) Experiment mit dem Computerprogramm »Eliza« eindrücklich gezeigt hat, das Menschen für einen einfühlsamen Psychiater halten, obgleich es nur simple Antworten schematisch generiert. Gleichwohl laufen die vielen vergeblichen Versuche, ›künstliche Intelligenz‹ (›KI‹) zu definieren, im Kern darauf hinaus, in heillos zirkulärer Weise ›KI‹ zu bestimmen als »making a machine behave in ways that would be called intelligent if a human were so behaving« (so schon McCarthy 1955) bzw. »… to create systems that are capable of performing tasks commonly thought to require intelligence« (Autorengruppe 2018). Mit diesen oder ähnlich unsinnigen Definitionen gelingt es nicht einmal, ›KI‹-Systeme von konventionellen Computersystemen zu unterscheiden. Deren Verhalten verkörpert, wie bei Maschinen immer, stets lediglich das durch die natürliche Intelligenz ihrer Konstrukteure erst gewonnene Wissen und Verständnis zugrunde liegender Sachverhalte (und kann natürliche Intelligenz folglich nicht ›erklären‹). Erzählungen über ›künstliche Intelligenz‹ sind mithin reine Selbsttäuschung.

So bleibt, um dieser Art Mystifikation von Computersystemen zu entkommen, nichts weiter übrig, als sich ihrer tatsächlichen Funktionsweise zu vergewis-

sern. Bekanntlich führen Computer nach dem ideellen Modell der Turingmaschine (Turing 1936) beliebige berechenbare Funktionen aus – und nichts sonst. Dazu besteht ihre Hardware aus binären Schaltsystemen (heute in Gestalt mikroelektronischer, hoch integrierter Prozessoren und Speicherchips hoher Taktfrequenz), auf deren Speicherzellen (›Registern‹) mittels elementarer Anweisungen programmgesteuert einfachste logische und arithmetische Operationen ausführbar sind. Benötigt werden dazu – basierend auf der theoretischen Grundlage der Booleschen Algebra – lediglich die Logikbausteine für Konjunktion, Disjunktion und Negation, die während eines Schalt-Takts auch die Addition beliebiger Dualzahlen realisieren (die Arithmetik mit Dualzahlen wird so auf logische Operationen der Booleschen Algebra zurückgeführt, veranschaulicht in Abb. 3). Auf dieser Basis lässt sich für die Ausführung beliebiger arithmetischer Operationen, letztlich auch mit negativen oder Gleitkomma-Zahlen, das anschließend beschriebene Rechnermodell einer minimalen abstrakten Maschine angeben, das diese Operationen durch wiederholte Ausführung nur ganz weniger elementarer Anweisungen ermöglicht. Zum Verständnis dieser Anweisungen ist noch wichtig, dass sich Logikbausteine in einem anfangs unbekannten Zustand befinden, der jeweils nur geändert werden kann (durch das Symbol := markiert) und daher vor Gebrauch erst in einen definierten Zustand versetzt werden muss.

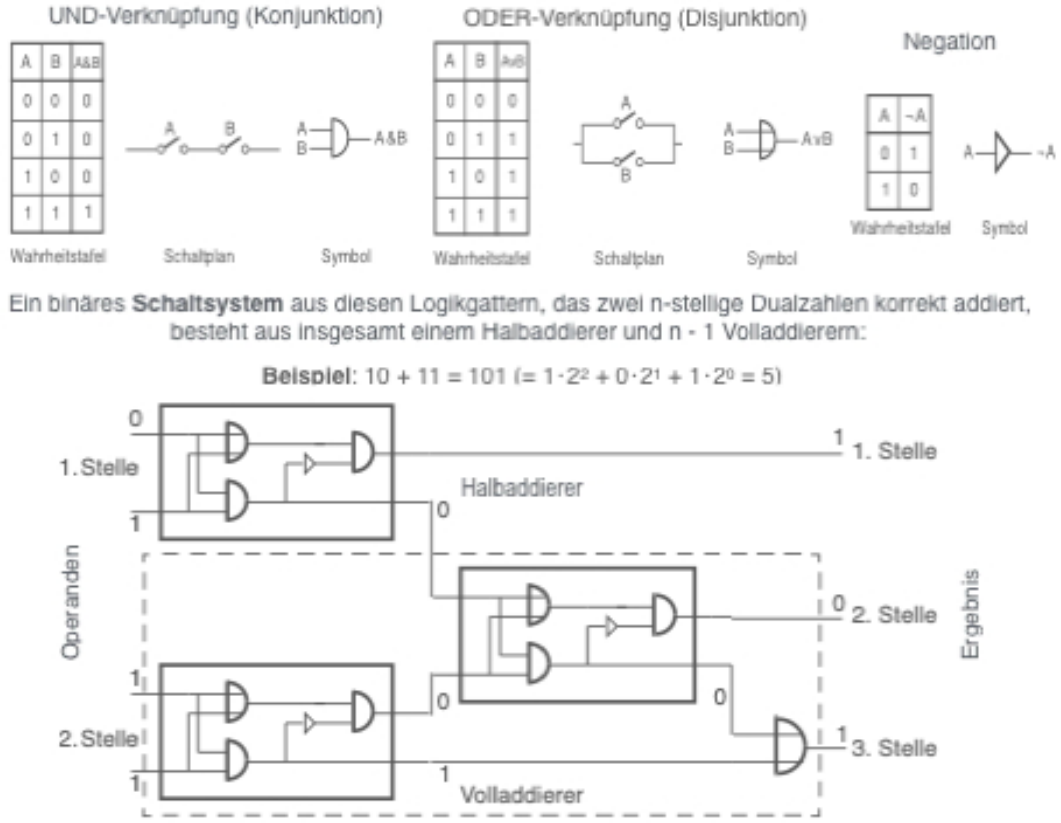


***Abb. 3**: Elementare Logikbausteine und damit realisiertes Schaltsystem zur Addition von Dualzahlen (Brödner 1997)*

Das Modell einer minimalen abstrakten Maschine ist gemäß Rekursionstheorie gleichmächtig wie die Turingmaschine, mithin wie diese zur Ausführung beliebiger berechenbarer Funktionen (Algorithmen) in der Lage (und setzt wie diese beliebig große Daten- und Programmspeicher für natürliche Zahlen inklusive der 0 voraus). Dazu genügen vier elementare Anweisungen:

$x := 0$	Löschen des Inhalts einer Speicherzelle,
$x := x + 1$	Erhöhen des Inhalts einer Speicherzelle um 1,
$x := x - 1$	Erniedrigen des Inhalts einer Speicherzelle um 1,
while $x \neq y$ do … end	Rekursion: Die While-Schleife wird solange durchlaufen bis $x = y$.

Damit lassen sich weitere, darauf aufbauende Anweisungen konstruieren; Beispiele:

- Laden einer Speicherzelle mit einer Konstanten n ($x :=$ n) durch Iteration: $x := 0$; $x := x + 1$; …; $x := x + 1$ (x wird n-mal um 1 erhöht),
- Zuweisen einer Variablen y an eine Variable x ($x := y$) mittels Rekursion: $x := 0$; while $x \neq y$ do $x := x + 1$ end,
- Erhöhen einer Variablen x um eine Variable y ($x := x + y$) mittels der Hilfsvariablen h durch Rekursion: $h := 0$; while $h \neq y$ do $x := x + 1$; $h := h + 1$ end,
- Addition ($z := x + y$) mittels Variablenzuweisung ($z := x$) und -erhöhung: ($z := z + y$),
- Boolesche und Relationsausdrücke lassen sich in arithmetische transformieren mit 1 für *true* und 0 für *false*.

Darin zeigt sich ein in der Computertechnik durchgängig genutztes Konstruktionsprinzip, das aus eindeutig bestimmten und vielfach erprobten einfachen Berechnungsfunktionen komplexere und reichhaltigere Funktionen zu bilden erlaubt (z.B. für Suchen, Sortieren, Bildverarbeitung oder Datentransfer u.v.a. mehr). Somit beruhen auch die komplexesten Programme und Softwaresysteme letztlich stets auf der Ausführung der hier beschriebenen Elementarfunktionen, allerdings operativ ausgeführt auf immer leistungsfähigeren Schaltsystemen, inzwischen in Milliarden Takten pro Sekunde. Das ist im Kern das ganze ›Geheimnis‹.

Bezeichnenderweise wurden fast alle dieser theoretisch-konzeptionellen Grundlagen der Computertechnik bereits im Zuge der Industrialisierung im 19. Jhdt. entwickelt, zuletzt im Verlauf der Grundlagenkrise der Mathematik in den 1930er Jahren, jedenfalls lange bevor die ersten Computer physisch realisiert werden konnten (vgl. nachstehende Übersicht). Daher ist auch die derzeit verbreitete Rede von der »vierten industriellen Revolution« geschichtsvergessener und ignoranter Unfug. Die seither erfolgte Entwicklung beruht denn auch weitgehend auf ständigen Verbesserungen der Hardware und im Software Engineering.

↑Inhalt↑

Historische Übersicht zu den Grundlagen der Computertechnik

1792-1801	Gaspard de Prony entwickelt und nutzt ein formularbasiertes Verfahren zur extrem arbeitsteiligen Neuberechnung mathematischer Tafeln (als Basis-Werkzeug für Ingenieurarbeit) im Dezimalsystem; das Formularschema der Rechenoperationen bildet die Urform eines Algorithmus (noch im WK II werden z.B. Tragwerke, V2-Flugbahnen u.v.a. so berechnet).
1805	Jacquard-Webstuhl, erste digital gesteuerte Arbeitsmaschine (mit Lochbrettern).
1812	Charles Babbage konzipiert die »Difference Engine« zur einfachen Berechnung von Polynomen: $f(x) = a_n x^n + \ldots + a_1 x + a_0$ (1822 prototypisch realisiert).
Um 1830:	Charles Babbage entwirft und programmiert die »Analytical Engine«; sie nimmt die von-Neumann-Architektur programmierbarer Universalrechner (Prozessor - Speicher - Steuerung) vorweg (scheitert aber an der Mechanik).
1847:	Die de Morganschen Gesetze $\neg(a \wedge b) = \neg a \vee \neg b$ und $\neg(a \vee b) = \neg a \wedge \neg b$ aufgreifend publiziert George Boole einen Logikkalkül (um 1888 von G. Peano als ›Boolesche Algebra‹ axiomatisiert); er bildet das logisch-funktionale Fundament für binäre Schaltsysteme (heutige Computer-Hardware, s. oben).
1860-1880:	C.S. Peirce entwickelt erstmals einen Prädikatenkalkül 1. Stufe, arbeitet an »logischen Maschinen« und entwickelt eine triadische Zeichentheorie, ohne die der allg. Computereinsatz nicht zu verstehen ist (äquivalent: G. Freges »Begriffsschrift« 1879; s. unten »algorithmisches Zeichen«)
1931:	Kurt Gödel beweist u.a. unter Verwendung rekursiver Funktionen die Unvollständigkeit formaler Systeme wie das der *principia mathematica* von B. Russell & A.N. Whitehead.
1936:	Alan Turing publiziert das ideelle Modell der ›Turingmaschine‹, definiert damit formal die Begriffe Algorithmus und berechenbare Funktion (äquivalent: λ-Kalkül von A. Church & S. Kleene).
1941/45:	Konrad Zuse entwickelt die Z3 als ersten binären Rechner (Relaistechnik) und den Plankalkül als erste (quasi-funktionale) Programmiersprache (markiert die Geburt des modernen Computers).

Damit stellt sich die entscheidende Frage, die letztlich über Wohl und Wehe des Einsatzes von Computersystemen bestimmt: Wie gelangt man von der sozialen Praxis in der Welt der Wissensarbeit und Wertschöpfung, ihren zumeist impliziten Handlungs- und Entscheidungsregeln, zu den Turing-berechenbaren bzw. partiell

rekursiven Funktionen, die allein auf binären Schaltsystemen ausführbar sind? Und wie gelangt man durch deren Anwendung wieder zurück zu Wissensarbeit höherer Leistung? Bei der Analyse dieser Übergangsprozesse stellt sich heraus, dass die vermeintliche »Macht der Algorithmen« tatsächlich die Macht der Interessen und Perspektiven ihrer Konstrukteure ist (bzw. der Einflüsse, denen diese ihrerseits im Kapitalverhältnis unterliegen).

Seit jeher erfordert nämlich die Gestaltung von Computersystemen für den praktischen Gebrauch eine sorgfältige und aufwendige Modellierung der stets zeichenbasierten sozialen Praktiken von kognitiver bzw. Wissensarbeit. Die der Wissensarbeit zugrunde liegenden Zeichenprozesse (»Semiosen«) müssen in einer zweckorientierten Perspektive sachgerecht modelliert werden durch partielle Explikation der tatsächlichen sozialen Praktiken in Gestalt begrifflichen Wissens über deren Gegenstände, Strukturen und Abläufe. Wesentliche theoretische Grundlage dafür ist die triadische Zeichentheorie von C.S. Peirce (1983), die Zeichen sozialer Praxis als dreistellige Relation ((R – O) – I) von physischem Zeichenkörper (»Repräsentamen« R), bezeichnetem Objekt (O) und Bedeutung zuweisendem Begriff (»Interpretant« I) auffasst. Als eigene soziale Praxis unterliegt auch die Modellbildung – die den eigentlichen Kern des Software Engineering ausmacht – stets den Interessen und Machteinflüssen beteiligter Akteure und durchläuft folgende Schritte der Abstraktion und Formalisierung (Andelfinger 1997):

- Semiotisierung: Die zunehmend präzisierte Beschreibung der betrachteten sozialen Praxis mittels Zeichen liefert ein perspektivisch reduziertes Abbild derselben als Ergebnis gemeinsamer Reflexion und Kommunikation der Akteure (Sprachanalyse, →›Ontologie‹). Ergebnis ist ein im wesentlichen sprachlich artikuliertes *Anwendungsmodell*.
- Formalisierung: Dessen Formalisierung durch Abstraktion von situations- und kontextgebundenen Bedeutungen mittels Verwendung standardisierter Zeichen und Operationen, durch Festlegung relevanter Bedingungen und Funktionen (funktionale Spezifikation) liefert ein *formales Modell*.
- Algorithmisierung: Die Überführung von Objekten, Abläufen und Funktionen des formalen Modells in auto-operational ausführbare Prozeduren in Form von Daten und berechenbaren Funktionen (Algorithmen) liefert schließlich das *Berechnungsmodell*.

Die berechenbaren Funktionen des in ausführbare Software transformierten Berechnungsmodells operieren auf Daten als auf rein syntaktische Aspekte reduzierten »Quasi-Zeichen« (Nöth 2002), deren binär kodierte »Repräsentamen« sie per Anweisung verändern, ohne Ansehen ihrer Bedeutung. Im Computersystem implementiert bilden sie »auto-operationale Formen« als Ausdruck abstrakter, formalisierter Handlungen (Floyd 2002). Auf diesem Wege der Modellierung und Formalisierung kann mittels Computersystemen »Kopfarbeit maschinisiert« werden (Nake 1992), wenigstens partiell (übrigens verhalten sich Menschen ebenfalls

wie eine Maschine, wenn sie ein Berechnungsverfahren wie etwa die schriftliche Division ausführen). Der Sinn der dabei ausgeführten »auto-operationalen Formen« muss freilich durch Aneignung für den praktisch wirksamen Gebrauch erst wieder erschlossen werden. So werden die Computersysteme durch Interpretation ihrer Funktionen im Handlungskontext der Arbeitspraxis, wieder in einen – freilich eben dadurch veränderten – Praxiszusammenhang gestellt (vgl. Abb. 4).

In der Interaktion mit Computern werden von Benutzern Zeichen (für Daten und damit operierende Funktionen) eingegeben, die für sie im jeweiligen Handlungskontext gewohnte Bedeutung tragen. Innerhalb des Computersystems werden diese außen sinnvoll interpretierbaren Zeichen auf binäre Signale als deren physischen Verkörperungen (R) reduziert, die mittels Programm nach vollständig festgelegten Anweisungen – dem Algorithmus – verarbeitet werden. Das mithin kausal determinierte Resultat dieser Signalverarbeitung kann dann bei Erscheinen an der Systemoberfläche erneut als Zeichen interpretiert werden. Über das mittels Kodierung beiden Zeichenprozessen gemeinsame Repräsentamen R auf der Benutzungsoberfläche sind diese fest gekoppelt. Intern verarbeitete Signale werden anstelle der intentionalen Interpretation durch Benutzer durch die Anweisungen des Algorithmus als kausalem Interpretanten determiniert, dessen Ergebnis R' das so kausal bestimmte Objekt bildet. Genau dies ermöglicht außen eine sinnvolle Interpretation (Abb. 4). So ist Interaktion mit Computern gekennzeichnet durch kausale Determination sinnfreier Signalverarbeitung im Innern und durch sinngebende Interpretation der an der Oberfläche als Zeichen gedeuteten Signale außerhalb. Der soziale Raum der Zeichenprozesse wird dabei nicht verlassen (weshalb Computer zurecht auch als semiotische Maschinen bezeichnet werden; Nake 2001).

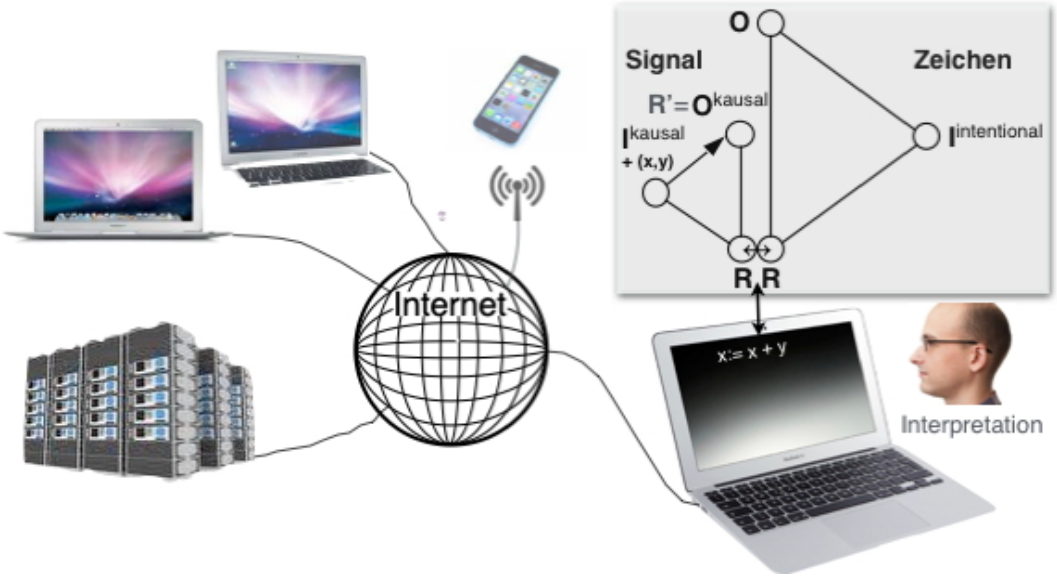


Abb. 4: *»Algorithmisches Zeichen« (Nake 2001): Als Einheit von Signal und Zeichen in vernetzten Computersystemen vermittelt es zwischen Signal und Sinn*

Damit wird auch die eigentlich neue Qualität deutlich: Verglichen mit bisherigen Medien handelt es sich bei global vernetzten Computersystemen um ein *instrumentelles Medium*, das in sich interaktiv nutzbare Werkzeuge zum Herstellen, Bearbeiten, Speichern und Auffinden beliebiger zeichenbasierter Gegenstände mit der Möglichkeit beliebigen Datentransfers über das Netzwerk zu anderen Teilnehmern vereint (Denning 2003). So können durch dieses integrierte Medium, im Zusammenspiel von Berechnung, Interaktion und Verbindung und unter Beachtung passender Standards und Handlungsvereinbarungen, weltweit räumlich verteilte, virtuelle Arbeitsräume computervermittelter Kooperation (CSCW; vgl. Schmidt 2011) sowie virtuelle Bibliotheken zur Nutzung kodifizierter Wissensbestände (WWW, verteilter Hypertext) geschaffen werden. Welchen gesellschaftlichen Nutzen das qualitativ neue Medium hat, entscheidet sich jedoch, wie bei anderen Medien auch, v.a. durch die Regeln, die sich eine Gesellschaft für den Umgang damit gibt.

Abschließend ist hier noch anzumerken, dass viele Computer nicht in Organisationen, sondern als Steuerungen für energetische, stoffliche oder logistische Prozesse eingesetzt werden. Mittels Sensoren zur Messung relevanter Prozessgrößen und Aktoren (Stellglieder) vermögen Steuerungen gezielt in Energie- und Stoffströme der Prozesse einzugreifen, aufgrund ihrer Algorithmen maschinelle Operationen zu koordinieren und damit den Prozessen ein zweckgemäß gewünschtes Verhalten zu verleihen, mit denen sie dann zu einem sog. »cyber-physischen System« (CPS) verschmelzen (daher werden sie auch »embedded systems« genannt). So können Steuerung und Prozess in gewünschter Weise zusammenwirken, um im Ergebnis Qualität und Mengenleistung der Prozesse zu steigern. Dies gelingt aber nur, weil von dem zu steuernden Prozess zuvor ein hinreichend genaues mathematisches oder heuristisches Modell in Zeichenform erstellt wurde, mittels dessen das Berechnungsmodell der Steuerung entworfen werden kann. Eben dies macht die Steuerung selbst ebenfalls zur semiotischen Maschine (sie verarbeitet Signale, aber keine Kräfte). Mithin können aus dieser Perspektive Computer auch als universelles Steuerungspotenzial verstanden werden, das jeweils per Programm als prozessspezifische Steuerung eingerichtet wird.

Nach diesen Einblicken in die tatsächliche Funktionsweise von Computern bleibt noch die Frage zu klären, was aus diesen Einsichten für die Erklärung des Produktivitätsparadoxons für eine künftig verbesserte Praxis geschlossen werden kann.

4 Schlussfolgerungen: Erklärung des Produktivitätsparadoxons und Perspektiven der Besserung

Wie aus der bisherigen Darstellung hervorgeht, wird mit Software massiv in soziale Praktiken organisierter Wissensarbeit interveniert. Durch Beobachtung und begriffliche Explikation vorgefundener Praktiken, v.a. der ihnen innewohnen-

den impliziten Regeln und Entscheidungslogik (›rationale‹) wird zunächst kodifiziertes Wissen über diese generiert. Auf dieser Basis kann zweckorientiert und interessengeleitet die zum Einsatz eines lauffähigen Softwaresystems führende Modellbildung und Formalisierung vorangetrieben und letztlich als nutzbares Computersystem implementiert werden. Um dessen Funktionen aber tatsächlich praktisch wirksam nutzen zu können, müssen diese freilich erst durch die Nutzer mühsam angeeignet werden, bevor sie routiniert verwendet werden können (vgl. Abb. 5). Bei komplexen Systemen, die verbundene Aufgaben vieler Wissensarbeiter betreffen, kann dieser Aufwand den für die eigentliche Entwicklung der Software bei weitem übersteigen (wie die einzelwirtschaftlich orientierte Empirie zum Paradoxon zeigt).

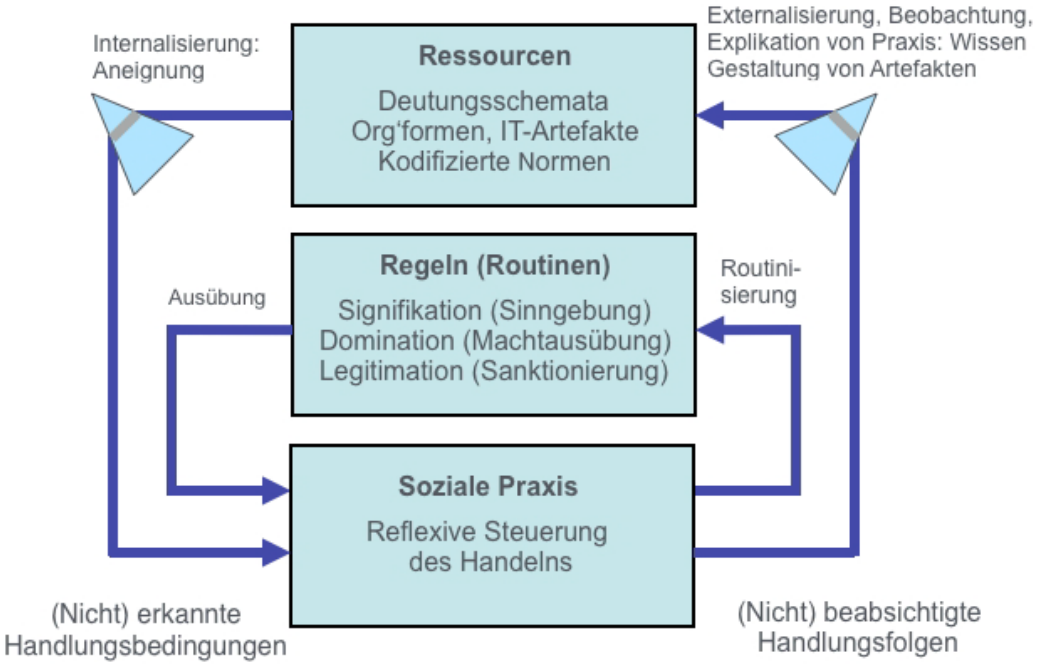


***Abb. 5**: Praxistheorie: Rekursive Konstitution von Handeln und Struktur (eigene Darstellung)*

Gestützt auf diese grundlegenden praxistheoretische Erkenntnisse (vgl. die Übersicht bei Reckwitz 2003; im einzelnen z.B. Orlikowski 2000), ist im vorliegenden Zusammenhang von großer Bedeutung zu erkennen, dass die Tätigkeiten, die im Kreislauf von Modellierung und Formalisierung, über den Einsatz als Computersystem bis zur Aneignung von dessen Funktionen vollzogen werden, insgesamt die organisationalen Praktiken stark verändern. Dabei ist zu reflektieren, dass sowohl Modellbildung und Implementierung des Computersystems als auch die Aneignung von dessen Funktionen für die Nutzung jeweils kreative, zweckorientierte und interessengeleitete Tätigkeiten sind, deren Verlauf und Ergebnis nicht vorherzusehen sind. Zudem ist dieser Kreislauf in hohem Maße *selbstbezüglich*, indem er maschinelle Funktionen für eine Praxis zu schaffen anstrebt, die am

Ende eine andere ist als die, für die sie konzipiert wurden. Soziale Praktiken der Wissensarbeit als Gegenstand der Modellierung geraten durch den Vorgang des Modellierens in Bewegung.

Damit erweist sich der Entwicklungsprozess der Software zugleich auch als Prozess der Organisationsentwicklung: Software ist Orgware, ein Medium des Organisierens. Dessen inhärente Selbstbezüglichkeit stellt die an der Entwicklung beteiligten und von ihr betroffenen Akteure vor enorme Herausforderungen. Zunächst lassen sich die Gebrauchstauglichkeit der Systemfunktionen wie auch die produktiven Wirkungen überhaupt nur im praktischen Einsatz, d.h. nach deren Aneignung für die eben dadurch aber veränderte Praxis der Wissensarbeit, angemessen beurteilen. Darüber hinaus gilt es im Verlauf des Prozesses möglichst früh die wechselseitige Ignoranz der Hauptakteure – das Unverständnis der Software-Ingenieure für die eigentlichen Aufgaben der Wissensarbeiter wie deren mangelndes Verständnis für die Möglichkeiten der Modellierung und Formalisierung – zu überwinden. Schließlich ist seitens des Projektmanagements dem Umstand Rechnung zu tragen, dass sich infolge der Selbstbezüglichkeit Anforderungen immer wieder ändern (Brödner 2008, Rohde et al. 2017).

So ergibt sich aus der semiotischen Natur von Computersystemen und ihrer Eigenschaft als Medium des Organisierens, dass Modellierung, Einführung oder auch Anpassung eines komplexen Softwaresystems nur gelingen können, wenn sie als integraler Teil eines umfassenderen Prozesses der Organisationsentwicklung verstanden und organisiert werden. Bislang noch zu oft praktizierte Vorgehensweisen, die sich sequentiell auf eine einmalige, umfassende Anforderungsanalyse mit nachfolgenden Entwicklungs- und Einführungsphasen gründen, sind dabei von vornherein zum Scheitern verurteilt. Vielmehr erfordert der Umstand der sozialen Einbettung semiotischer Maschinen ein *partizipativ* und *reflexiv* angelegtes, *evolutionäres Vorgehen* mit kurzen, überschaubaren Revisionsschleifen, in dem wiederholt Zyklen der Anforderungsanalyse, Softwareentwicklung, Implementation, Erprobung sowie der *formativen Evaluation* erreichter Resultate durchlaufen werden. In einer solchen Entwicklungsspirale werden die Systemfunktionalität und deren Aneignung im Rahmen restrukturierter Prozesse in jeweils kleinen, bewusst begrenzten Entwicklungszyklen hervorgebracht und so wiederholt der Bewertung und Einflussnahme unterworfen. So bleiben Anforderungen und Entwicklungsaufgaben der einzelnen Zyklen überschaubar und halten die Risiken des Scheiterns in Grenzen (ggf. ist nur der letzte Zyklus zu revidieren). Dabei sind bewährte Methoden des Software Engineering zu nutzen, reichen aber allein nicht aus. Vielmehr müssen sich die Beteiligten dabei insgesamt über alle Aspekte ihrer im Entstehen begriffenen neuen sozialen Struktur verständigen, was insbesondere die produktive Verbindung verschiedener Sichtweisen, die Bewältigung von Konflikten und den Ausgleich unterschiedlicher Interessen einschließt.

Dazu ist es erforderlich, die wechselseitige Ignoranz der Akteure zu überwinden, indem die praktische Erfahrung aus der wirklichen Arbeit, bereits expliziertes Prozesswissen und das Wissen über technische Möglichkeiten und Grenzen von Computersystemen produktiv miteinander verbunden werden. Dazu bedarf es hinreichender Analyse der Praxis, etwa mit ethnografischen Methoden, und der Initiierung eines Kommunikations- und Lernprozesses, in dem die Beteiligten unterschiedliche Perspektiven zusammenführen, ein geteiltes Verständnis des Arbeitsprozesses entwickeln und dabei eine geteilte ›Sprache der Artefakte‹ nutzen. Darin lassen sich explizites Wissen über den sich ändernden Arbeitsprozess und die darin zu nutzenden Softwareartefakte artikulieren und gemeinsam reflektieren. Dabei müssen sich die Akteure zunächst auch über Ziele, Aufgaben und Grundsätze ihrer künftigen Zusammenarbeit – über ihre Geschäftsstrategie – sowie über dazu passende Strukturen und Abläufe der Arbeits- und Wertschöpfungsprozesse verständigen, um daraus abzuleiten, wie diese durch Programme und Daten wirksam unterstützt werden können. Wer Softwaresysteme adäquat gestalten und produktiv nutzen will, muss Organisationsentwicklung betreiben, unter Beteiligung aller betroffenen Akteure von Beginn an.

Vor diesem Hintergrund ist auch leicht nachzuvollziehen, dass sich derzeit sog. ›agile Methoden‹ des Software Engineering (vgl. Hanser 2010) zunehmend verbreiten, denen ein derartiges zyklisch evolutionäres Entwicklungsmuster zugrunde liegt und – sofern sie der dargelegten Herausforderungen eingedenk auch konsequent umgesetzt werden – auch beträchtliche Erfolge verzeichnen. Allerdings wird diese notwendige Bedingung oft nicht eingehalten und durch Rückfall in alte gewohnte Praktiken werden mögliche Erfolge verschenkt. Allerdings stellen die einschneidenden, aber notwendigen Veränderungen das Management vor ein schwer zu bewältigendes Dilemma: Derart weit reichende kollektive Lernprozesse sind grundsätzlich nach Verlauf und Ergebnis offen und bedeuten daher aus herkömmlicher Managementsicht einen kaum erträglichen Kontrollverlust. Andererseits sind sie den hier angestellten Überlegungen zufolge notwendige – wenngleich auch nicht hinreichende – Bedingung für den Projekterfolg. So ist traditionell denkendes Management hin- und hergerissen zwischen dem Risiko eines mit hoher Wahrscheinlichkeit scheiternden Organisations-Entwicklungsprojektes und der Befürchtung eines riskanten Kontrollverlusts über den Projektverlauf. Genau hierin ist auch der Grund auszumachen, warum – trotz wiederkehrender Erfahrungen des Scheiterns – so wenige Manager wagen, Projekte zur Organisationsentwicklung reflexiv und evolutionär anzugehen. Das Dilemma wird jedoch drastisch gemildert, wenn man die Prozedur einer zyklischen formativen Projektevaluation ernst nimmt und sich auf diese neue Form der Kontrolle durch reflexive Steuerung einlässt.

Aus der sozialen Einbettung von Softwaresystemen und deren engem Verwobensein mit den sozialen Praktiken von Organisationen ergibt sich ferner, dass

Produktivitätsfortschritte nicht aus erhöhter Leistungsfähigkeit der Hardware, sondern letztlich vor allem aus der Reorganisation der zugrunde liegenden Geschäftsprozesse und der dabei benötigten Wissensarbeit entstehen können. Eben hierin unterscheiden sich semiotische Maschinen grundsätzlich von klassischen Maschinen der Energie- und Stoffumwandlung, die zwecks Produktivitätssteigerung zusätzlich Naturkräfte und -effekte erschließen und nutzen können.

Selbst wenn viele einzelne Aufgaben der Wissensarbeit durch Implementierung umfangreicher Berechnungsverfahren – etwa bei der Berechnung der Festigkeit oder Dynamik mechanischer Strukturen (FEM), der Steuerung von komplexen logistischen Prozessen oder der Simulation von Wettervorgängen – weitgehend automatisiert werden können, wird die dadurch eingesparte Arbeit durch Rebound-Effekte vermehrter oder vielfacher Anwendung zumeist wieder kompensiert. Nicht selten stehen aber auch algorithmisch festgelegte Prozeduren bei der Bewältigung von unvorhergesehenen oder Ausnahmesituationen eher als Hindernis im Wege und erfordern zusätzliche Arbeit sie zu umgehen. Zudem ist zu berücksichtigen, dass sowohl die Modellierung und Formalisierung bis hin zur Konstruktion der Algorithmen der Berechnungsmodelle als auch die Aneignung von deren Funktionen für die praktisch wirksame Nutzung und Interpretation beträchtlichen Aufwand erfordern. Gerade in diesen arbeitsaufwendigen Prozessen wird über Erfolg oder Misserfolg des Einsatzes von Softwaresystemen entschieden. Umfangreiche zusätzliche Wissensarbeit wird darüber hinaus auch für die fortlaufende Entwicklung und Vermittlung von Methoden des Software Engineering und der Berechnungsverfahren benötigt. Zusammengenommen erklärt das recht gut das Ausbleiben nennenswerter gesamtwirtschaftlicher Produktivitätseffekte.

Schließlich wird damit auch verständlich, warum das unreflektierte Gerede von ›Digitalisierung‹, insbesondere von ›künstlicher Intelligenz‹ und ›maschinellem Lernen‹, die eigentlichen Herausforderungen weithin ignoriert und die wirklichen Vorgänge beim Einsatz fortschrittlicher Computersysteme nicht angemessen beschreibt. Die verwendeten Metaphern führen direkt ins Reich der Mythen und verbergen – wie beim trickbasierten Zaubern – die eigentlich banale Funktionsweise auf doppelte Weise: das, was versteckt wird, und den Vorgang des Versteckens selbst. So verbergen diese Metaphern, dass die ganze Intelligenz jeweils in den Tätigkeiten der Modellierung und Formalisierung sozialer Praxis und der Aneignung der Softwarefunktionen für den praktisch wirksamen Gebrauch steckt, während das Computersystem lediglich die Algorithmen des so gewonnenen Berechnungsmodells ausführt. So laufen gebräuchliche Verfahren ›maschinellen Lernens‹ (etwa Verfahren mit Stützvektoren oder Entscheidungsbäumen, künstliche neuronale Netze, K-Means Clustering oder lineare Regression) im Kern auf bloße Funktions-Approximation an vorhandene Daten hinaus (das hat mit ›Lernen‹ im herkömmlichen Sinn nichts zu tun). Dabei werden meist alt bekannte mathematische Methoden, ausgeführt als eine Art ›Hochgeschindigkeits-

Statistik‹, auf extrem schneller Hardware eingesetzt – allerdings mit stets nur wahrscheinlichen und daher unsicheren Ergebnissen, zudem oftmals auf Basis fragwürdiger Qualität der zur Anpassung benutzten Daten. Zugleich bleibt der Einsatz der Verfahren auf den Problemtyp beschränkt, für den sie geschaffen sind. Mithin sind diese Metaphern ein irreführender Etikettenschwindel.

Der Erfolg des Einsatzes fortschrittlicher Computersysteme hängt folglich entscheidend von der analytischen und methodischen Kompetenz der Entwickler und dem Verständnis der Nutzer ab, sie für ihre Wissensarbeit sinnvoll zu verwenden. Für die künftige Entwicklung dieser Systeme ist dementsprechend ein grundlegender Perspektivwechsel angesagt: von einem daten- und methodengetriebenen Vorgehen zu problemorientierter Herangehensweise. Statt zu fragen, wie oft nur zufällig verfügbare Daten und Methoden der Modellierung sozialer Praxis auf unterschiedliche Aufgaben der Wissensarbeit angewandt werden können, um sie durch Automaten zu ersetzen, ist es weit zielführender, praktisch relevante Probleme der Steigerung von Flexibilität und Produktivität von Wissensarbeit organisatorisch anzugehen, sie bei der Bewältigung ggf. auch, aber nicht nur, mit methodisch sorgfältig entwickelten Berechnungsmodellen effektiv zu unterstützen. Erforderlich ist nichts weniger als ein Wandel im Verständnis und der Perspektive von Entwicklung und Gebrauch der Computertechnik, der Perspektivwechsel vom Automatisierungsmittel zum Medium des Organisierens, methodisch von der Systemgestaltung zur Strukturation sozialer Praxis. Allein auf diesem Wege können Computersysteme auch künftig zur effektiven Bewältigung gesellschaftlicher Herausforderungen beitragen (z.B. zu sozial-ökologischer Transformation). Das belegt jedenfalls die ganze bisherige Entwicklung, sowohl durch die Erfolge bei der Steigerung der Leistung computerunterstützter Wissensarbeit als auch durch die Sackgassen misslungener Automatisierung, während entgegen stehende Behauptungen lediglich dem Fetisch des Computers als ›intelligenter Maschine‹ huldigen.

5 Literatur

Accenture (2015): Economic Literature Review: Impact of Technology on Productivity/Labor Productivity, https://pdfs.semanticscholar.org/1c85/219b3f962cb8f9884ec93a4c8e2556ed5fc7.pdf

Acemoglu, D.; Autor, D.; Dorn, D.; Hanson G.H. & Price, B. (2014): Return of the Solow Paradox? IT, Productivity, and Employment in US Manufacturing, American Economic Review: Papers & Proceedings 104 (5), 394–399

Andelfinger, U. (1997): Diskursive Anforderungsanalyse. Ein Beitrag zum Reduktionsproblem bei Systementwicklungen in der Informatik, Frankfurt/M: Peter Lang

Anders, G. (1973): Die Antiquiertheit des Menschen. Bd. 1, München: Beck

Autorengruppe (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, Oxford: Future of Humanity Institute u.a. 02/2018, https://arxiv.org/pdf/1802.07228.pdf

Babbage, C. (1832): Die Ökonomie der Maschine, Nachdruck der Originalübersetzung von 1833, hg. von P. Brödner, Berlin: Kulturverlag Kadmos 1999

Boes, A.; Kämpf, T.; Langes, B. & Lühr, T. (2014): Informatisierung und neue Entwicklungstendenzen von Arbeit, Arbeits- und Industriesoziologische Studien 7 (1), 5-23

Bresnahan, D.F.; Brynjolfsson, E. & Hitt, L.M. (2002): Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-level Evidence, The Quarterly Journal of Economics 117 (1), 339–376

Brödner, P. (2008): Das Elend computerunterstützter Organisationen, in: D. Gumm (Hg.): Mensch – Technik – Ärger? Zur Beherrschbarkeit soziotechnischer Dynamik aus transdisziplinärer Sicht, Münster: Lit-Verlag, 39-60

Brödner, P. (1997): Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen, Berlin: edition sigma

Brynjolfsson, E. (2003): The IT Productivity Gap, Optimize, Issue 21, July 2003

Brynjolfsson, E. (1993): The Productivity Paradox of Information Technology, CACM 36 (12), 67-77

Brynjolfsson, E. & Hitt, L.M. (2000): Beyond Computation: Information Technology, Organizational Transformation and Business Performance, Journal of Economic Perspectives 14 (4), 23-48

Dedrick, J.; Gurbaxani, V. & Kraemer, K.L. (2003): Information Technology and Economic Performance: A critical review of the empirical evidence, ACM Computing Surveys 35 (1), 1-28

Denning, P.J. (2003): Great Principles of Computing, CACM 44 (11), 15-20

Emam, K.E. & Koru, A.G. (2008): A Replicated Survey of IT Software Project Failures, IEEE Software 25 (5), 84-90

Floyd, C. (2002): Developing and Embedding Autooperational Form, in: Y. Dittrich; C. Floyd & R.

Klischewski (Eds.) (2002): Social Thinking – Software Practice, Cambridge (MA): MIT Press, 5-28

Gordon, R.J. (2016): The Rise and Fall of American Growth. The U.S. Standard of Living since the Civil War, Princeton: Princeton University Press

Gordon R.J. (2014): The Demise of U.S. Economic Growth: Restatement, Rebuttal, and Reflections, NBER Paper

Haug, W. F. (2005): Vorlesungen zur Einführung ins »Kapital«, Hamburg: Argument

Hanser, E. (2010): Agile Prozesse: Von XP über Scrum bis MAP, Berlin Heidelberg: Springer

Jorgenson, D.W.; Ho, M.S. & Stiroh, K.J. (2008): A Retrospective Look at the U.S. Productivity Growth Resurgence, Journal of Economic Perspectives 22 (1), 3-24)

Kretschmer, T. (2012): Information and Communication Technologies and Productivity Growth: A Survey of the Literature, OECD Digital Economy Papers, No. 195, Paris: OECD Publishing

Landauer, T.K. (1995): The Trouble with Computers. Usefulness, Usability, and Productivity, Cambridge (MA): MIT Press

Lutz, B. (1987): Das Ende des Technikdeterminismus und die Folgen. Soziologische Technikforschung vor neuen Aufgaben und neuen Problemen, in: Lutz, B. (Hg.):Technik und sozialer Wandel. Verhandlungen des 23. Deutschen Soziologentages in Hamburg 1986, Frankfurt/M: Campus Maucher, I. (2001): Komplexitätsbewältigung durch Entwicklung und Gestaltung von Organisation München: Hampp

Maucher, I. (Hg.) (1998): Wandel der Leitbilder zur Entwicklung und Nutzung von PPS-Systemen. München: Hampp

McKinsey Global Institute (2002): How IT Enables Productivity Growth. The US experience across three sectors in the 1990s, San Francisco: MGI

Nake, F. (2001): Das algorithmische Zeichen, in: W. Bauknecht; W. Brauer & T Mück (Hg.): Informatik 2001. Tagungsband der GI/OCG Jahrestagung, 736-742

Nake, F. (1992): Informatik und Maschinisierung von Kopfarbeit, in: W. Coy; F. Nake; J.-M. Pflüger; A.

Rolf; J. Seetzen; D. Siefkes & R. Stransfeld (Hg.): Sichtweisen der Informatik, Braunschweig Wiesbaden: Vieweg, 181-201

Naur, P. & Randell, B. (eds.) (1968): Software Engineering. Report on a conference sponsored by the NATO Science Committee, Brussels: NATO

Nöth, W. (2002): Semiotic Machines, Cybernetics and Human Knowing 9 (1), 5-22

Orlikowski, W. J., 2000: Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations, Organization Science 11(4), 404-428

Peirce, C.S. (1983): Phänomen und Logik der Zeichen, Frankfurt/M: Suhrkamp

Potthoff, I. (1998): Empirische Studien zum wirtschaftlichen Erfolg der Informationsverarbeitung. In: Wirtschaftsinformatik 40, 54-65

Reckwitz, A. (2003): Grundelemente einer Theorie sozialer Praktiken. Eine sozialtheoretische Perspektive, Zeitschrift für Soziologie 32 (4), 282-301

Rohde, M.; Brödner, P.; Stevens, G.; Betz, M. & Wulf, V. (2017): Grounded Design – a Praxeological IS Research Perspective, Journal of Information Technology 32 (2), 163-179

Ropohl, Günter (1991): Technologische Aufklärung. Beiträge zur Technikphilosophie, Frankfurt/M: Suhrkamp.

Schmidt, K. (2011): Cooperative Work and Coordinative Practices. Contributions to the Conceptual Foundations of Computer-Supported Cooperative Work (CSCW), London: Springer

Smith, A. (1776): Der Wohlstand der Nationen – eine Untersuchung seiner Natur und seiner Ursachen, aus dem Engl. übertragen u. mit einer Würdigung von H.C. Recktenwald, München: Beck 1974

Standish Group International (2016): CHAOS Report 2015, https://www.standishgroup.com/sample_research_files/CHAOSReport2015-Final.pdf

Stiroh, K. J. (2002). Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say? American Economic Review 92 (5), 1559-1576

Solow, R.M. (1987): We'd Better Watch Out, New York Times Book Review, July 12

Turing, A.M. (1950): Computing Machinery and Intelligence, Mind 49, 433-460

Turing, A.M. (1936): On Computable Numbers. With an Application to the Entscheidungsproblem, in: M. Davis (Ed.): The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems, and Computable Functions, New York 1965, 116-151

Verein Deutscher Ingenieure (Hg.) (1991): Technikbewertung – Begriffe und Grundlagen. Erläuterungen und Hinweise zur VDI-Richtlinie 3780, VDI Report 15, Düsseldorf: VDI.

Weber, M. (1921). Wirtschaft und Gesellschaft. Tübingen: Mohr Siebeck 1976

Weizenbaum, J. (1966): ELIZA – A Computer Program for the Study of Natural Language Communication Between Man and Machine, CACM 9 (1), 36-45

↑Inhalt↑

›Informatik‹ – eine Wissenschaft auf Abwegen

Zusammenfassung: Das Produktivitätsparadoxon der Computertechnik und vergleichsweise häufiges Scheitern großer Softwareprojekte deuten an, dass inadäquate Begriffe und Beschreibungen falsche Vorstellungen und Erwartungen über Funktionsweisen computertechnischer Artefakte hervorrufen. Gestützt auf einen elaborierten Zeichenbegriff werden fachliche Grundlagen kritisch beleuchtet im Hinblick darauf, wie die sich von klassischen Maschinen fundamental unterscheidenden Artefakte im Zusammenhang mit der Organisation von Wissensarbeit zu verstehen sind. Ferner werden die Verwendung irreführender Metaphern sowie daraus resultierende Illusionen über Nutzen und Gefahren der Computertechnik mit der Folge beträchtlicher Fehlallokation von Ressourcen analysiert.

Abstract: The productivity paradox of computer technology as well as relatively frequent failures of huge software projects indicate that inadequate concepts and descriptions evoke erroneous ideas and expectations about modes of operation of computer artifacts. Based on an elaborated sign concept, basic principles of computer technology are being analysed with respect to how these artifacts, fundamentally different from classical machines, can be apprehended in connection with the organization of knowledge work. Moreover, the use of deceptive metaphors and resulting illusions about benefits and risks of computer technology, as well as resulting misallocation of resources, are being contemplated.

Schlüsselbegriffe: Informationsverarbeitung, Zeichen, soziotechnische Systeme, maschinelles Lernen

1 Einführung: Uneingelöste Versprechen

Seit rund fünf Dekaden werden Computer in entwickelten Gesellschaften für vielfältige Aufgaben der Wissensarbeit auf breiter Front eingesetzt und weltumspannend vernetzt. Trotz massiver Investitionen in Computertechnik, die diejenigen in Produktionstechnik längst übersteigen, werden die damit verbundenen hohen Erwartungen an die Steigerung der Arbeitsproduktivität immer wieder enttäuscht (sog. ›Produktivitätsparadoxon‹ der Computertechnik; vgl. Gordon 2014, Brödner 2020, Horn 2021). Zudem zeigt sich stets aufs Neue, dass große Software-Projekte trotz hoher Budgetüberschreitungen die angestrebte Funktionalität oftmals verfehlen oder sogar ganz scheitern (Standish Group 2015). Derart beispiellose, in anderen Technikfeldern so kaum zu verzeichnende Krisenerscheinungen, insbesondere die Diskrepanz zwischen ausbleibenden Produktivi-

tätsgewinnen und gleichwohl aufrecht erhaltenen hohen Erwartungen, lassen dahinter systematische Gründe vermuten. Unter den mutmaßlichen Gründen für diese Krisenanzeichen stechen tief sitzende, sich immer wieder neu bildende Missverständnisse der wahren Natur der Computertechnik, ihrer Funktionsweise und Potenziale hervor.

Diese Missverständnisse zeigen sich auch in unterschiedlichen Sichtweisen der Wissenschaftsdisziplin ›Informatik‹, darunter in zwei dominanten: zum einen in derjenigen, welche Computer als »informationsverarbeitende Maschinen« versteht, zum anderen derjenigen, deren Grundverständnis sich im sog. »computational model of the mind« widerspiegelt. Beide sind eng miteinander verwandt und verlangen nach kritischer Analyse ihrer von Beginn an nicht wirklich geklärten, z.T. auch widersprüchlich verwendeten Grundbegriffe (z.B. maschinelle ›Verarbeitung von Information‹ versus ›Datenverarbeitung‹, Verhältnis von ›Software‹ zu ›Hardware‹). Desgleichen verdient auch der irreführende metaphorische Gebrauch von Begriffen aus fachfremden Wissensdomänen (»intelligente Agenten«, »maschinelles Lernen«, »künstliche neuronale Netze (KNN)«, »autonome Systeme« etc.) eine kritische Auseinandersetzung. Darin offenbart sich eine beträchtliche Unsicherheit über die Natur des Fachs und seiner Gegenstände. Sie bescheren dem Fach zwar immer wieder große öffentliche Aufmerksamkeit, tragen aber auch wesentlich zu Missverständnissen seiner Artefakte und deren Funktionsweise bei. Selbst nach einem halben Jahrhundert stürmischer Entwicklung dieser Disziplin, sind trotz ihrer großen wirtschaftlicher Bedeutung noch immer grundlegende Fragen nach dem eigentlichen Untersuchungsgegenstand, nach ihm angemessenen Methoden und dessen besonderem Charakter offen.

Zu beträchtlichen Teilen resultieren diese Schwierigkeiten aus dem Umstand, dass vernetzte Computersysteme als Artefakte zur maschinellen Verarbeitung, Speicherung und Übertragung von Daten tiefgreifend in soziale Praktiken der Kommunikation und Kooperation intervenieren, aus denen die Daten stammen und die ihnen Bedeutung verleihen. Bei Gestaltung und Gebrauch dieser Artefakte sind daher Erkenntnisse aus ganz unterschiedlichen Wissensdomänen gleichermaßen gefragt: Logik bzw. Arithmetik, Physik und Sozialwissenschaft. Bei der soziotechnischen Gestaltung (vgl. z.B. Mumford 2006) von Computersystemen und ihrer produktiven Verwendung müssen folglich Erkenntnisse aus allen diesen Wissensdomänen sinnvoll zusammengeführt werden: Wie lassen sich Aspekte sozialer Kommunikation und Kooperation in Form einer Abfolge logischer und arithmetischer Operationen modellieren und formalisieren, wie lassen sich diese maschinell ausführen und wie die damit erzielten Ergebnisse wieder interpretieren und sinnvoll nutzen?

Diesen Fragen wird hier nachgegangen mit dem Ziel, zu ihrer Klärung und zum besseren Verständnis der technischen Herausforderungen beizutragen. Dazu wird zunächst ›Information‹ als dem Fach ›Informatik‹ zugeschriebener Grundbe-

griff kritisch beleuchtet und als ungeeignet für die Darstellung wesentlicher Vorgänge der Computertechnik ausgewiesen. Mit dem triadischen Zeichenbegriff, wie er im wesentlichen von den Logikern C.S. Peirce und G. Frege elaboriert wurde, wird dann eine alternative begriffliche Grundlage vorgestellt, die es insbesondere erlaubt, logisch-arithmetische Operationen und ihre maschinelle Ausführung durch binäre Schaltsysteme mit deren Interpretation im Kontext sozialer Praktiken in Verbindung zu setzen. Eben das damit explizit beschreibbare Zusammenspiel maschineller Operationen auf bedeutungslosen Alphabetzeichen mit deren sinnhaftem Gebrauch in sozialer Praxis lässt die Besonderheiten der Computertechnik und der mit ihrem Einsatz auftretenden Probleme besser verstehen. Damit können dann schließlich auch die vielen irreführend antropomorphisierenden Erzählungen entkräftet werden, die sich um die Computertechnik ranken.

2 ›Information‹: Ein schillernder und trügerischer Begriff

Seit Claude Shannon (1948) seine Analyse nachrichtentechnischer Übertragungssysteme als »A Mathematical Theory of Communication« publizierte (›Nachrichten‹ werden dabei ausdrücklich als reine Wortfolgen ohne ›Bedeutung‹ betrachtet) und darin, gestützt auf seine Vorläufer Nyquist und Hartley, ein rein syntaktisch über einem endlichen Alphabet definiertes Maß für die Menge an durch einen Nachrichtenkanal übertragener »Information« definierte, geistert dieses Wort durch unterschiedlichste wissenschaftliche Diskurse; so wurde es sogleich etwa auch im einflussreichen Buch »Kybernetik« von Norbert Wiener (1948) aufgegriffen und erreichte damit später Disziplinen übergreifende Aufmerksamkeit. Aufgrund rein formaler Ähnlichkeit dieser für Signale geltenden Definition mit derjenigen der physikalischen Entropie hat Shannon dieses Maß, seiner Sache offenbar selbst nicht sicher, alternativ auch »Entropie« genannt.

Die darin angelegte Begriffsverwirrung vergrößert sich noch mit dem kurz darauf unter dem leicht veränderten Titel »The Mathematical Theory of Communication« publizierten Nachdruck der Shannonschen Arbeit, ergänzt um eine äußerst fragwürdige, von positivistisch inspiriertem Zeitgeist geprägte Arbeit von Warren Weaver, dem langjährigen (bis 1955 amtierenden) Direktor für Naturwissenschaften der Rockefeller Foundation als einflussreicher Institution der Wissenschaftsförderung. Hierin bringt Weaver, anders als Shannon, auch die semantischen und pragmatischen Aspekte von Nachrichten wieder ins Spiel; so behauptet Weaver etwa, die technisch effiziente Signalübertragung bestimme auch Fragen des Verstehens und der Anerkennung einer Nachricht: „... a larger part of the significance comes from the fact that the analysis at Level A [signal transmission, PB] discloses that this level overlaps the other levels more than one could possibly naively suspect. Thus the theory of Level A is, at least to a significant

degree, also a theory of levels B [semantics, PB] and C [pragmatics, PB]“ (Shannon & Weaver 1949: 3).

Die begriffliche Wirrnis wird noch vertieft durch die Alltagssprache, in der unter »Information« gemeinhin der Inhalt einer Mitteilung, Auskunft oder Belehrung als Ergebnis ihrer Interpretation verstanden wird. Und aus sozialwissenschaftlicher Perspektive ist für das Handeln im Rahmen einer sozialen Praxis ohnehin nur die kontextabhängige Bedeutung einer Nachricht wesentlich, wie es in der prägnanten Definition: »Information ist jeder Unterschied, der etwas ausmacht« (»any difference that makes a difference«, Bateson 1980: 250) markant zum Ausdruck kommt. Somit existieren unter derselben Benennung ›Information‹ mindestens drei höchst unterschiedliche, untereinander unverträgliche Begriffe nebeneinander. Unglücklicherweise werden davon aber im Kontext soziotechnischer Systemgestaltung computerunterstützter Arbeit zumindest zwei, der rein syntaktische von Shannon und der sozialwissenschaftliche Begriff, zugleich gebraucht, um soziale Praktiken im Umgang mit Computertechnik und deren innere Funktionsweise angemessen beschreiben zu können (Brödner 2016). So kann es auf die dabei zentralen Fragen: Womit genau Computer eigentlich operieren – mit ›Information‹ oder Signalen? und wie sich das im Arbeitsprozess genau auswirkt? keine klaren Antworten geben, zumal dabei auch der alltagssprachliche Begriff stets mitspielt. Bis heute sind denn auch alle weiteren Versuche, zu einem Disziplinen übergreifend einheitlichen Informationsbegriff zu gelangen, kläglich gescheitert (vgl. etwa den in den Heften 4 und 5 des Informatik Spektrums 26 (2003) geführten Diskurs sowie Janich 2006).

Gleichwohl gilt ›Information‹ als Grundbegriff der für Computertechnik zuständigen Fachdisziplin ›Informatik‹ – eine erstmals 1957 vom Nachrichteningenieur Karl Steinbuch geprägte, 1962 von Philippe Dreyfus in Frankreich eingeführte, 1967 von der Academie Française als »Wissenschaft von der rationalen, insbesondere maschinellen Verarbeitung von Information« geadelte Wortschöpfung, für die der Begriff bereits konstitutiv ist und die »automatische Informationsverarbeitung« durch Computer zum grundlegenden Paradigma erhebt (vgl. etwa das Positionspapier »Was ist Informatik?« der GI (2006) oder einschlägige Lehrbücher, z.B. Bauer & Goos 1971). Dabei fällt freilich auf, dass außer in kurzen einleitenden, wenig erhellenden Abschnitten über ›Information‹ in den für das Fach wesentlichen Darstellungen der theoretischen Informatik (Berechenbarkeit), der technischen Informatik (Schaltsystementwurf) und der praktischen Informatik (Softwaretechnik) der Begriff ›Information‹, offenbar verzichtbar, nirgends mehr vorkommt. Er ist ja auch weder ein Begriff der Physik (in keinem deutsch- oder englischsprachigen Handbuch der Physik kommt er als Stichwort vor), noch ein Begriff der Logik oder Mathematik, die beide wesentliche Grundlagen des Fachs bereitstellen (im Englischen heißt das Fach denn auch »computer« bzw. genauer: »computing science«). Dieser bereits in den Grundlagen des Fachs angelegte gedankliche Wirrwarr trägt beträchtlich zu den verbreiteten Missverständnissen

rund um die Computertechnik bei, indem der irreführende Informationsbegriff, aufgeladen durch soziale Bedeutung suggerierende Konnotationen, ständig dazu verführt, maschinell vollzogene Zustandsänderungen bedeutungsloser Signale mit deren sinnhafter Interpretation im Kontext sozialer Praktiken zu verwechseln. Das wird durch eine Reihe weiterer durchweg anthropomorphisierender Fehlbenennungen noch gesteigert (vgl. Brödner 2021 und Abschnitt 4). Vorderhand wird mithilfe eines elaborierten Zeichenbegriffs eine alternative begriffliche Grundlage gelegt.

3 Logik der Zeichen: Eine alternative begriffliche Grundlage der Computertechnik

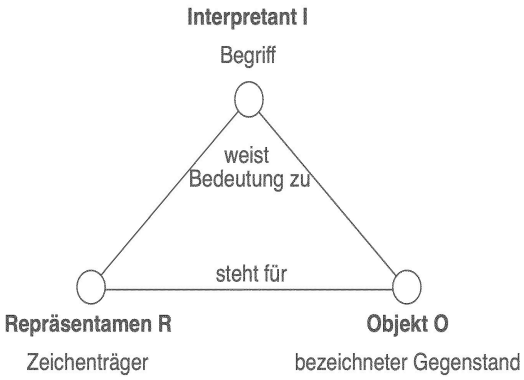
Mit der Entstehung eines gewaltigen gesellschaftlichen Mehrprodukts in den fruchtbaren Schwemmland-Gebieten der alten Flusskulturen wurden vor gut 5000 Jahren Schrift- und Zahlzeichen entwickelt. Mit deren Hilfe konnten Vorgänge der Verwaltung vielfältiger und umfangreicher Vorräte – neben solchen der Landvermessung (›Geometrie‹), der Ertrags- und Steuerberechnung – genau beschrieben und ›buchführend‹ symbolisch nachvollzogen werden. Sehr viel ältere Praktiken des Zählens wurden dafür mittels arithmetischer Operationen zu Berechnungsverfahren fortentwickelt (so stammen etwa früheste bekannte Berechnungen der Kreiszahl oder einer rekursiven Funktion aus diesem Kontext; vgl. Ifrah 1989). Der springende Punkt dabei ist die Repräsentation und Beschreibung der materiellen Vorgänge durch einzelne diskrete Zeichen eines endlichen Alphabets, deren Verknüpfung zu Worten und Manipulation nach bestimmten Regeln, die bis heute gelten. So unterscheiden sich etwa dazu entwickelte Berechnungsverfahren der Buchführung nicht grundsätzlich von denen moderner ERP-Systeme, außer dass letztere umfassender und vielseitiger sind und v.a. maschinell ausgeführt werden. Im weiteren geschichtlichen Verlauf werden zudem durch radikale Abstraktion von den Inhalten schriftsprachlicher Aussagen logische Aussageverknüpfungen formalisiert: So publiziert George Boole, die ›de Morganschen Gesetze‹

$$\neg(a \wedge b) = \neg a \vee \neg b \quad \text{und} \quad \neg(a \vee b) = \neg a \wedge \neg b$$

aufnehmend, Mitte des 19. Jahrhunderts einen Logikkalkül, der, um 1888 von Peano als ›Boolesche Algebra‹ axiomatisiert, schließlich das logisch-funktionale Fundament für binäre Schaltsysteme als heutiger Computer-Hardware bildet.

In diesem Kontext erarbeitet der amerikanische Logiker C.S. Peirce erstmals einen Prädikatenkalkül 1. Stufe, arbeitet an »logischen Maschinen« und entwikkelt dabei eine triadische Zeichentheorie**,** die heute allgemein als logische Grundlage für den Einsatz von Computern gelten kann (als äquivalent ist die gleichzeitig und unabhängig davon entwickelte Begriffsschrift von Gottlob Frege anzusehen).

Der Peircesche triadische Zeichenbegriff (Peirce 1983) bestimmt ein Zeichen allgemein als dreistellige Relation aus



1. einem *Repräsentamen R,* dem physischen Zeichenkörper oder -träger (ein als Zeichen gedeuteter Gegenstand oder Vorgang, ›Alphabetzeichen‹)**,**
2. einem *Objekt O*, dem bezeichneten Gegenstand oder Vorgang und
3. einem *Interpretanten I*, der Bedeutung, die ein Interpret dem Paar (*R, O*) situativ und kontextabhängig zuschreibt (Begriff).

Somit ist ein Zeichen »etwas, das für jemanden in einer bestimmten Hinsicht oder Fähigkeit für etwas steht«. Der Begriff ist rekursiv, denn der Interpretant ist selbst ein Zeichen, das bezeichnet und interpretiert werden kann. Dabei beschreibt (vgl. z.B. IFIP 1998)

- *Syntaktik* die auf *R* reduzierten Aspekte von Zeichenprozessen (*womit* wird bezeichnet? → Daten, repräsentiert durch Signale)**;**
- *Semantik* deren auf die Relation (*R – O*) reduzierten Aspekte (*was* wird bezeichnet? → Information i. S. von ›steht für‹)**;**
- *Pragmatik* die sozio-kulturell anerkannt wirksamen Aspekte von Zeichenprozessen (*wie* wird Bezeichnetes in einem Handlungskontext gedeutet? → begriffliches Wissen).

In dieser »praxistheoretischen« Perspektive (Reckwitz 2003) werden Gegenstände und Tatsachen (›Fakten‹) der sozialen Welt – sog. ›institutionelle Tatsachen‹ – erst durch Kommunikation und Kooperation als Formen zeichenbasierten koordinierten Handelns aufgrund »geteilter Intentionalität« (Tomasello 2009) geschaffen und, insoweit anerkannt, auch erhalten: »Wir sorgen dafür, dass etwas der Fall ist, indem wir es als etwas repräsentieren, was der Fall ist« (Searle 2012). Als Paradebeispiel dafür kann die ausschließlich mit gedachten (daher empirisch nicht zugänglichen), durch Alphabetzeichen und formale Regeln geschaffenen Objekten operierende Mathematik gelten, die Hilbert zufolge »ein Spiel mit wenigen Regeln und bedeutungslosen Zeichen auf Papier« ist (dafür bieten etwa die der Computertechnik zugrunde liegenden Theorien der Berechenbarkeit oder der Automaten und formalen Sprachen reiches Anschauungsmaterial).

Grundsätzlich können beliebige physische Gegenstände oder Vorgänge als Repräsentamen fungieren und als Zeichen für etwas gedeutet werden. Mit dem triadischen Zeichenbegriff wird zudem der zentralen Tatsache Rechnung getragen, dass ›Bedeutung‹ keine natürliche Eigenschaft physischer Gegenstände oder Vorgänge ist, sondern diesen in einem konkreten Handlungskontext sozialer Praktiken, in bewusster Interaktion durch sinngebende Interpretation ihrer Funktionen, zugeschrieben wird; sie ist stets subjektiv und kann daher nicht Gegenstand

wiederholbarer, maschineller Operationen sein. Mit diesem Zeichenbegriff lässt sich jedoch eine logische Verbindung herstellen zwischen der physischen Welt kausal determinierter Wirkungen und der sozialen Welt der Genese intentional bedingter, situativ zugeschriebener Bedeutungen. Damit ist er in besonderem Maße geeignet, Vorgänge der sozialen Praxis computerunterstützter Wissensarbeit und ihrer soziotechnischen Gestaltung präzise zu beschreiben und zu erklären (vgl. Abb. 1).

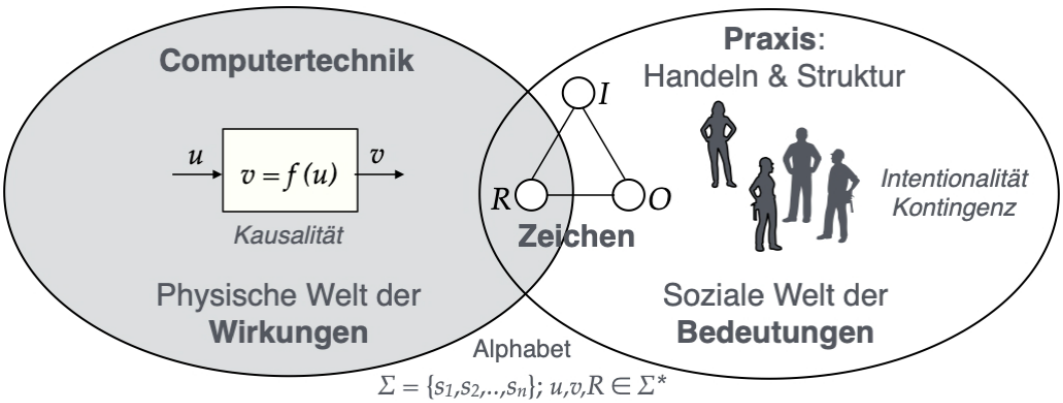


Abb. 1: Das triadische Zeichen als logische Verbindung von Physik und Semantik (eigene Darstellung)

Der Computereinsatz beginnt mit der stets partiellen Modellierung und Formalisierung bestimmter Aspekte einer sozialen Praxis: Dabei dienen durch Selbst- und Fremdbeobachtung mittels Sprache, in Gestalt von Text oder Grafik, vereinbarte teils deskriptive, teils präskriptive Modelle gewünschter sozialer Praktiken als Grundlage, um daraus streng formale Beschreibungen in Form von Algorithmen und Datenstrukturen abzuleiten. Auf diese Weise werden bestimmte Aspekte einer zeichenbasierten sozialen Interaktion beschreibende, aber perspektivisch reduzierte Modelle in Programme als formaler Darstellung von Berechnungsmodellen abgebildet: $v = f(u)$; $v, u \in \Sigma^*$. Durch diese Reduktion gehen infolge der Abstraktion sowohl Bezeichnungen $(R - O)$ als auch Bedeutungen $I - (R - O)$ verloren; die per Programm fremdgesteuerte, allein mit $R \in \Sigma^*$ operierende Maschine ›Computer‹ ›weiß‹ nicht mehr, was sie prozessiert und warum. Programme werden per Compiler automatisch in ausführbaren Maschinencode übersetzt. Während ihrer Laufzeit existiert die Maschine ›Computer‹ nur noch als reine ›Hardware‹, als Abfolge von Signalzuständen binärer Schaltsysteme, zweckmäßig gesteuert durch den Maschinencode. Vom Computer für bestimmte zulässige Eingaben per Programm erzeugte Ergebnisse können dann aufgrund von Aneignung und Kenntnis der Eigenschaften des Berechnungsmodells f im Kontext der Verwendung interpretiert und erneut mit Bedeutung aufgeladen werden. Insgesamt wird dadurch die soziale Praxis restrukturiert.

Software hat somit zwei Seiten: Programme werden in der sozialen Welt der Bedeutungen nach deren Anforderungen und Zwecken gestaltet und vermögen – nach automatischer Compilierung in Maschinencode – als solche in der physischen Welt der Wirkungen Signal-Zustandsänderungen in Schaltsystemen zu steuern. Eben diese Doppelnatur verführt zur verbreiteten Illusion vermeintlich kognitiver Leistungen: In der sozialen Welt der Bedeutungen bilden Programme und Daten eine formale Beschreibung maschineller Funktionen als (freilich oft nur schwer) lesbarem Text. Software gehört in dieser Hinsicht zur sozialen Welt der Bedeutungen, sie beschreibt, was ablaufen soll und wozu. In Form compilierter Programme ist sie aber zugleich auch funktionale Steuerung einer Maschine. Daher muss sie interessenabhängigen sozialen Zwecken und Anforderungen genügen und zugleich eine formal korrekte Beschreibung berechenbarer Funktionen (Algorithmen, Berechnungsmodell) liefern.

Durch die automatische Verwandlung in maschinell ausführbaren Code und die dabei insgesamt vollzogene Abstraktion und formale Reduktion auf logisch determinierte Signalzustände verlieren Programme jeglichen Bezug zur sozialen Welt von Bezeichnungen und Bedeutungen. Der Code steuert dann nur Zustandsänderungen bedeutungsloser Signale in binären Schaltsystemen und wird so zum Herzstück der Maschine ›Computer‹ in der physischen Welt der Wirkungen, ohne noch zu ›wissen‹ wozu. Erst das damit erzeugte Ergebnis ist aufgrund der Kenntnis der Funktionen zugrunde liegender Berechnungsmodelle in der sozialen Welt wieder interpretierbar.

Diese Betrachtungen gelten im übrigen *mutatis mutandis* auch für den Fall des Computer Einsatzes zur Steuerung physischer Prozesse (sog. »cyber-physische Systeme«): In diesem Fall ist die vorgängig ebenfalls in der sozialen Welt der Bedeutungen vorgenommene mathematische oder doch zumindest heuristisch begründete formale Modellierung des betrachteten physischen Prozesses zwingende Voraussetzung dafür, überhaupt ein formal notiertes Programm zur Steuerung des physischen Prozesses entwerfen zu können, die diesem dann in kompilierter Form, durch Mess- und Stellsignale vermittelt, ein zweckgemäßes Verhalten zu verleihen vermag. So beruht auch dieser Steuerungsentwurf auf einem zeichenbasierten Prozessmodell. Im derart durch Signale vermittelten Zusammenspiel von Steuerung und Prozess bilden dann beide eine selbsttätig prozessierende Einheit in der physischen Welt der Wirkungen, deren Funktionsweise und Güte aber wiederum nur in der sozialen Welt der Bedeutungen beurteilt werden kann.

Damit erweisen sich Computersysteme als technische Artefakte, die ausschließlich berechenbare Funktionen vollziehen, die ihrerseits während des Entwurfs aus Zeichenprozessen (›Semiosen‹) kooperativer Wissensarbeit (›Kopfarbeit‹) zweckgemäß abstrahiert und als »autooperationale Formen« (Floyd 2002) verwirklicht werden. Deren maschinelle Ausführung besteht ausschließlich aus physischen Vorgängen einer entsprechend programmierten Steuerung von Signalzuständen in Schaltsystemen. Somit bewegt sich die soziotechnische Gestal-

tung computerunterstützter Arbeit, verstanden als Anpassung von Form, Funktion und Handlungskontext, stets im Spannungsfeld des technisch Machbaren, der Formbarkeit von Natur bzw. sozialer Praxis, und des sozial Wünschenswerten, abhängig von jeweiligen Interessen und Machtverhältnissen.

Diese Sicht der Dinge spiegelte sich denn auch lange Zeit in sachlich angemessenen Bezeichnungen für die so verstandenen technischen Artefakte, etwa »elektronischen Rechenanlagen« zur »Datenverarbeitung« oder »Prozessrechner« zur Prozesssteuerung, bevor widersinnig die Rede von »Informationsverarbeitung« erneut Oberhand gewann. Dementsprechend wird auch in internationalen Begriffsnormen sehr klar zwischen »Information« (als »knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning«) und »Daten« (als »reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing«) als logischer Form physischer Signale unterschieden (vgl. ISO/IEC 2015).

4 Wortspiele: Irreführende Benennungen verführen zu Selbsttäuschung

Aus den dargelegten Zusammenhängen geht insbesondere hervor, dass es, um Computertechnik überhaupt zweckmäßig nutzen zu können, zunächst erforderlich ist, durch Analyse, Modellierung und Formalisierung von Vorgängen sozialer Praktiken (bzw. physischer Systeme im Falle von Prozesssteuerung) berechenbare Funktionen zu identifizieren, mit dem Ziel, sie zu aufgabenangemessen gestalteten, interaktiv nutzbaren Berechnungsverfahren (Algorithmen) zu bündeln, die auch maschinell ausführbar sind. Diese zwar erprobten, aber aufgrund der nötigen Interpretationsleistungen nur mit Nutzerbeteiligung zu bewältigenden, meist sehr aufwändigen und von Rückschlägen bedrohten Entwicklungsprozesse soziotechnischer Gestaltung werden neuerdings mit sog. »KI«-Methoden zu umgehen versucht (vgl. Brödner 2020).

Technische Grundlage derartiger Umgehungsversuche ist die Verfügbarkeit sehr großer Datenbestände (»Big Data«) und extrem gesteigerter Rechenleistung. Zusammen genommen ermöglichen sie, sog. adaptive Systeme zu gestalten, deren Funktionsweise einfach auf vorgegebenen generischen, komplex strukturierten Funktionen (etwa in Gestalt von KNN oder Entscheidungsbäumen) beruht. Sie enthalten eine große Zahl von Parametern, die mittels – größtenteils alt bekannten – Optimierungsverfahren an die Vielzahl von außen aufgenommener Daten optimal angepasst (»trainiert«) werden; mathematisch handelt es sich um Verfahren der Funktions-Approximation. Mittels der so »trainierten« Funktionen lassen sich dann für eingegebene weitere Daten zugehörige Resultate errechnen, z.B. in Aufgaben der Klassifikation oder Entscheidung.

Mit diesem vermeintlichen Befreiungsschlag von den Mühen der Modellierung und Formalisierung handelt man sich freilich sogleich neue, noch schwerwiegendere Probleme ein. Diese Vorgehensweise

- beruht auf theorielosem Probieren mit unterschiedlich vorstrukturierten Funktionen, damit allein auf dem Erfahrungswissen der Entwickler hinsichtlich ihrer Eignung und Reichweite (Abhängigkeit von Erfahrung);
- steht und fällt mit der Qualität der Daten hinsichtlich Erhebungsmethode, Korrektheit oder Verzerrungen, die aber meist höchst fragwürdig und zumindest im vorhinein kaum einschätzbar ist (unsichere Datenqualität);
- setzt in Vergangenheit angepasstes Verhalten auch in Zukunft fort und negiert damit die prinzipielle Kontingenz sozialer Praktiken (Behinderung von Wandel);
- macht das Zustandekommen der zudem stets mit Unsicherheit behafteten Ergebnisse für die Nutzer gänzlich intransparent und behindert damit im Gebrauch deren zielführendes instrumentelles Handeln (Handlungshindernis).

Diese im Grundsatz simple, methodisch jedoch höchst fragwürdige, letztlich auf schierer Rechenleistung zur Verarbeitung großer Datenmengen beruhende Vorgehensweise wird nun durch eine Fülle aufregender, aber irreführender Fehlbenennungen wie »künstliche Intelligenz«, »Maschinelles Lernen«, »KNN« oder gar »autonome Systeme« zu verschleiern versucht. Damit wird nahtlos an eine verbreitete Praxis missbräuchlicher Verwendung von Metaphern angeknüpft, die von Beginn an die ›Kybernetik‹ und Computertechnik begleiten. So ist etwa, von damals vorherrschenden positivistischen und behavioristischen Vorstellungen verführt, schon früh von »Elektronengehirnen«, »Giant Brains or Machines That Think« *(*Buchtitel von E.C. Berkeley 1949) die Rede. Derartige Metaphern rufen gänzlich falsche Vorstellungen funktionaler Vergleichbarkeit lebendiger, sich selbst organisierender (›autopoietischer‹) Gehirne mit Computern als toten, bewusst hergestellten, Natureffekte nutzenden semiotischen Artefakten hervor, haben sich aber sogleich im bis heute dominanten sog. ›computational model of the mind‹ der ›Kognitionswissenschaften‹ niedergeschlagen. So behauptet etwa Steinbuch in dafür typischer reduktionistischer Manier, »dass zur Erklärung geistiger Funktionen höchstwahrscheinlich keine Voraussetzungen gemacht werden müssen, die über die Physik hinausgehen« (Steinbuch 1965: 2). In diesem Zusammenhang erscheint ferner auch zur Beantwortung der Frage, ob Computer denken, der Turing-Test (Turing 1950) als sinnlos, weil er – ganz im Geiste des Bahaviorismus – aus äußerlich beobachtetem Verhalten irrtümlich auf deren innere Funktionsweise zu schließen sucht.

Denken lässt sich nicht mit Berechnen gleichsetzen oder nachahmen. Computer ›denken‹ nicht, sondern vollziehen nur, wie in Abschnitt 3 gezeigt, von Menschen ausgedachte und vereinbarte Algorithmen. Nicht ihrer Funktionsweise als Maschine, sondern der kreativen Leistung ihrer Entwickler wohnt »Intelligenz« inne. Intuition, »Abduktion« (Peirce 1934: CP 5.189), Begreifen und Ur-

teilen sind menschliche kognitive Leistungen, die über maschinelles Berechnen und Schließen hinausgehen, wie etwa die Beweisführung zu den Gödelschen Unvollständigkeitssätzen eindrücklich aufzeigt: Im Kern beruht der Beweis gerade darauf, dass er als kompetenter Mathematiker metamathematische Prädikate über Terme eines formalen Systems, z.B. die Beweisbarkeit, per operativer Codierung als Formeln im System selbst auszudrücken vermag. Damit gelingt ihm, eine Formel im System so zu konstruieren, dass sie über sich als einem durch ihn als wahr erkannten Satz aussagt, nicht beweisbar zu sein (vgl. Gödel 1931).

Als gestaltete Artefakte funktionieren Computersysteme auch keineswegs »autonom«, wie das derzeit ständig behauptet wird, sondern allenfalls ›automatisch‹. Tatsächlich sind Systeme nur dann ›autonom‹, wenn sie sich die Regeln ihres Verhaltens selbst zu setzen vermögen; Computersysteme sind im Gegensatz dazu ›heteronom‹, ihr Verhalten ist durch fremdbestimmte Algorithmen determiniert, sie sind selbsttätig, aber nicht *von selbst* tätig, schon gar nicht setzen sie die Regeln ihres Funktionierens selbst. Ihnen ist folglich auch keinerlei ›Handlungsträgerschaft‹ eigen, wie von sozialwissenschaftlicher Seite gelegentlich behauptet wird (vgl. z.B. Rammert 2003), denn diese würde das Vermögen voraussetzen, eigene Ziele setzen zu können. Stattdessen ergibt sich auch das zielführende Verhalten von ähnlich irreführend als »intelligente Agenten« (Russell & Norvig 2009) bezeichneten adaptiven Systemen aus den von ihren Schöpfern durch jeweilige ›Verlustfunktionen‹ und die darauf fußenden Algorithmen zur Funktions-Approximation implizit vorgegebenen Zielen.

Besonderes Unheil scheint indes die Benennung »Maschinelles Lernen« für Berechnungsverfahren anzurichten, wie sie in adaptiven Systemen in Gestalt von KNN, insbesondere Verfahren des »Deep Learning«, oder Entscheidungsbäumen derzeit häufig zum Einsatz kommen (und oft viel Aufsehen erregen). Solche Verfahren haben mit herkömmlichen Begriffen von ›Lernen‹ im Sinne des Gewinnens von Einsichten nichts zu tun, sondern dienen allein der zweckgemäßen Anpassung des Systemverhaltens an Daten von außen (mit allen genannten, damit einhergehenden Problemen). Gleichwohl führt das ohne Umschweife zu den in zahlreichen sozialwissenschaftlichen Publikationen und gesellschaftlichen Diskursen Platz greifenden Missverständnissen von Computersystemen als »lernenden Maschinen« oder gar »lernenden Algorithmen« (als Gipfel grotesker Selbsttäuschung, denn Algorithmen zeichnen sich gerade dadurch aus, dass ihr Verhalten vollständig und eindeutig festgelegt ist). Daher sind es auch nicht, wie oft befürchtet, irgendwelche ›Algorithmen, die Macht über deren Nutzer‹ gewinnen, sondern die den Entwicklern und ihren Auftraggebern eigenen Machtpotenziale und Interessenlagen bestimmen, wie sie diese mittels entsprechend gestalteter Computerartefakte über deren Nutzer ausüben.

Diese gravierenden Missverständnisse der Computertechnik bilden Wegmarken wachsenden Realitätsverlustes. An dieser kritischen Sicht der Dinge vermag auch der blendend schöne Schein von Berichten über aufsehenerregende Er-

folge bei einzelnen spezifischen Aufgaben, etwa der maschinelle Sieg über den weltbesten Go-Spieler, beeindruckende maschinelle Sprach-»Übersetzungen« oder Bildklassifikationen, nichts zu ändern. Wie ein raffiniert konstruierter hydraulischer Schaufelbagger menschliche Muskelkraft in bestimmten mechanischen Bewegungsabläufen um Größenordnungen zu übertreffen vermag, kann auch heuristisch scharfsinnig genutzte, schiere maschinelle Rechenleistung menschliche kognitive Leistungen bei bestimmten genau definierbaren Aufgaben weit übersteigen (zu technischen Details vgl. Brödner 2019). Für diese bestimmten Aufgaben spezifisch konstruierte Berechnungsverfahren sind aber nicht, jedenfalls nicht ohne großen Aufwand, auf andere Aufgaben übertragbar.

6 Schlussfolgerung: Wissenschaft als Ideologie

Alles in allem zeigt sich, dass – leider auch von Fachvertretern propagierte – irreführende Benennungen und falsche Metaphern für Artefakte der Computertechnik leider dauerhaft üblich geworden sind und von Beginn an zu immer wieder neuen erheblichen Missverständnissen ihrer genauen Funktionsweise wie auch zur Verschleierung der eigentlichen Probleme ihrer Gestaltung und ihres Gebrauchs führen. Durch diese mittels begrifflichen ›Framings‹ induzierte Mystifizierung von Computerartefakten werden diesen stets aufs Neue kognitive Fähigkeiten wie ›Intelligenz‹, ›Lernen‹, ›Entscheiden‹, ›Verstehen‹ oder gar ›Autonomie‹ angedichtet. Diese Art der Mythenbildung führt freilich im Weiteren zu beträchtlichen Folgeschäden, indem sie Diskurse fehlorientiert und gesellschaftliche Ressourcen in hohem Maße fehlalloziert (obgleich dabei einzelne Akteure stark profitieren: Cloud-Anbieter an der Rechenleistung, Netzbetreiber an der Bandbreite, Wissenschaftler an Forschungsmitteln, Finanzkapital an Spekulationsgewinnen etc.). An der auf diese Weise vorherrschenden Mystifizierung von Computerartefakten erweist sich einmal mehr, wie auch »Technik und Wissenschaft [zu] Ideologie« (Habermas 1968) mutieren können.

Indes zeichnet sich mit der Häufung realer Probleme bei Entwicklung und Einsatz datengetriebener adaptiver Systeme eine zunehmend kritischere Sicht auf den derzeitigen Stand der Technik ab. So hat erst kürzlich Wolfgang Wahlster (2020), ein führender Vertreter der »KI«-Forschung, resümiert: »*Inzwischen wurde die Bedeutung von kausalem Hintergrundwissen in der jeweiligen Anwendungsdomäne klar. Ohne Kausalwissen können von datengetriebenen Lernverfahren abgeleitete Scheinkorrelationen kaum als Unsinn erkannt und aussortiert werden. Auch die derzeit auf der ganzen Welt geforderte Erklärungsfähigkeit von KI-Systemen kann ohne explizite Modelle, wie sie etwa seit jeher für die Beschreibung von Naturgesetzen verwendet werden, kaum in einer für den Menschen nachvollziehbaren Weise erreicht werden.*» Freilich bleibt dabei ausgeblendet, dass eben solche Anstrengungen zur Realisierung wissensbasierter Systeme der sog. »symbolischen KI« der 1980er Jahre an den Hürden nur sehr

begrenzter Explizierbarkeit impliziten Wissens und dessen situativer Anwendung gescheitert sind. Insgesamt sind die Mystifizierung von Computerartefakten und die damit verbundenen Fehlorientierungen untrügliche Anzeichen einer Krise des Fachs ›Informatik‹, die auch zu einem beträchtlichen Teil das eingangs zitierte Produktivitätsparadoxon zu erklären vermögen. Rückbesinnung auf fachspezifische Grundlagen und Klärung relevanter Grundbegriffe sowie sachgerechte Bezeichnungen und Vorgehensweisen sind daher dringend geboten.

Der Orientierung dieser Rückbesinnung in der Perspektive soziotechnischer Gestaltung computerunterstützter Wissensarbeit (vgl. als jüngstes Bsp. APRODI-Verbund 2021) mögen folgende abschließende Überlegungen dienen: ›Informatik‹ (eigentlich: Computertechnik) ist eine Ingenieurwissenschaft, die technische Artefakte zur gebrauchstauglichen maschinellen Verarbeitung von Signalen (logisch: ›Daten‹ – nicht ›Information‹) in Zeichenprozessen sozialer Praxis analysiert, gestaltet und bewertet. Von früheren Ingenieurdisziplinen (Maschinenbau, Elektrotechnik, chemische oder biologische Verfahrenstechnik) unterscheidet sie sich dadurch grundlegend, dass nicht allein Naturprozesse, sondern v.a. auch durch Zeichen vermittelte Prozesse sozialer Interaktion (Kommunikation und Kooperation) analysiert und in Form von berechenbaren Funktionen modelliert werden. Diese Besonderheit erfordert auch besondere Vorgehensweisen und Methoden.

Ein Kennzeichen der Praxis sozialer Interaktion ist es, dass sie ›Objektivität‹ nur insoweit gewinnt, als die Gemeinschaft der beteiligten sprach- und handlungsfähigen Akteure sie als gemeinsame Wirklichkeit erlebt. Notwendige Bedingung dafür ist, dass sich die kommunikativ handelnden Akteure über das verständigen, was in ihrer geteilten Welt bzw. der Wirklichkeit ihrer sozialen Praxis der Fall ist und in ihr bewirkt werden soll. Als in einer sozialen Praxis mittels Zeichen kommunikativ handelnde Akteure müssen Informatiker und Anwender technischer Artefakte maschineller Datenverarbeitung folglich eine gemeinsame Wirklichkeit mit geteilten Interpretationsschemata für ihre Gegenstände und die Art und Weise, sie zu gebrauchen, entwickeln. Zum Kernbestand ›informatischer‹ Bildung muss daher neben der Gestaltung und Bewertung datenverarbeitender Artefakte auch die Befähigung zum systematischen Dialog mit anderen Wissenschaftlern, Anwendern und Nutzern dieser Artefakte gehören.

7 Literatur

APRODI-Verbund (2021): Betriebliche Digitalisierung erfolgreich gestalten, Eschborn: RKW Rationalisierungs- und Kompetenzzentrum

Bateson, G. (1980): Mind and Nature. A Necessary Unity, Toronto: Bantam Books

Bauer, F.L & Goos, G. (1971): Informatik. Eine einführende Übersicht, 2 Bde., Berlin Heidelberg: Springer

Brödner, P. (2021) »Machines that think« – die »KI«-Illusion und ihre Wurzeln, in: Klaus Lenk & Jörg Pohle (Hg.): Der Weg in die »Digitalisierung« der Gesellschaft – Was können wir aus der Geschichte der Informatik lernen? Marburg: Metropolis, 67-82

Brödner, P. (2020): Das Produktivitätsparadoxon der Computertechnik, in: H.J. Bontrup & J. Daub (Hg.): Digitalisierung und Technik – Fortschritt oder Fluch? Perspektiven der Produktivkraftentwicklung im modernen Kapitalismus, Köln, PapyRossa, 114-144

Brödner, P. (2019): Grenzen und Widersprüche der Entwicklung und Anwendung ›Autonomer Systeme‹, in: H. Hirsch-Kreinsen & A. Karacic (Hg.): Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt, Bielefeld: transcript 2019, 69-97

Brödner, P. (2016): Verwirrung durch »Information«? Zur Kritik des Paradigmas »maschineller Informationsverarbeitung«, in: F. Fuchs-Kittowski & W. Kriesel (Hg.): Informatik und Gesellschaft. Festschrift zum 80. Geburtstag von Klaus Fuchs-Kittowski, Frankfurt/M: Peter Lang, 297-308

Floyd, C. (2002): Developing and Embedding Autooperational Form, in: Dittrich, Y.; Floyd, C.; Klischewski, R. (Eds.): Social Thinking – Software Practice, Cambridge (MA): MIT Press, 5-28

Gesellschaft für Informatik (2006): Was ist Informatik? Unser Positionspapier, https://gi.de/fileadmin/GI/Hauptseite/Themen/was-ist-informatik-lang.pdf

Gödel, K. (1931): Über formal unentscheidbare Sätze der principia mathematica und verwandter Systeme I, Monatshefte für Mathematik und Physik 38, 173-198

Gordon, R.J. (2014): The Demise of U. S. Economic Growth: Restatement, Rebuttal, and Reflections, NBER Paper, https://content.csbs.utah.edu/~mli/Economics%207004/Gordon_NBER%20P383F%20Sequel_140126.pdf

Habermas, J. (1968): Technik und Wissenschaft als Ideologie, Frankfurt/M: Suhrkamp

Horn, A. (2021): Zehn Jahre Industrie 4.0 und kein Produktivitätsfortschritt, Novo Argumente, https://www.novo-argumente.com/artikel/zehn_jahre_industrie_4.0_und_kein_produktivitaetsfortschritt

IFIP (1998); FRISCO Report: A Framework of Information Systems Concepts, https://www.mathematik.uni-marburg.de/~hesse/papers/fri-full.pdf

Ifrah, G. (1989): Universalgeschichte der Zahlen, Frankfurt New York: Campus ISO/IEC 2382 (2015): Information Technology – Vocabulary, https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en

Janich, P. (2006): Was ist Information? Frankfurt/M: Suhrkamp

Mumford, E. (2006): The Story of Socio-technical Design. Reflections on its successes, failures and potential, in: Information Systems Journal 16 (4), 317-342

Peirce, C.S. (1983): Phänomen und Logik der Zeichen, Frankfurt/M: Suhrkamp

Peirce, C.S. (1934): Collected Papers Vol. 5, Pragmatism and Pragmaticism, Cambridge (MA): Harvard University Press

Rammert, W. (2003): Technik in Aktion, in: T. Christaller & J. Wehner (Hg.): Autonome Maschinen – Perspektiven einer neuen Technikgeneration, Wiesbaden: Westdeutscher Verlag, 289-315

Reckwitz, A. (2003): Grundelemente einer Theorie sozialer Praktiken. Eine sozialtheoretische Perspektive, Zeitschrift für Soziologie 32 (4), 282-301

Russell, S. & Norvig, P. (2009): Artificial Intelligence: A Modern Approach, 3rd. ed., Essex (GB): Pearson

Searle, J.R. (2012): Wie wir die soziale Welt machen, Berlin: Suhrkamp

Shannon, C. (1948): A Mathematical Theory of Communication, The Bell System Technical Journal 27, July, 379–423 & October, 623–656

Shannon, C. & Weaver, W. (1949): The Mathematical Theory of Communication, Urbana: University of Illinois Press

Standish Group (2015): CHAOS Report 2015, https://standishgroup.com/sample_research_files/CHAOSReport2015-Final.pdf

Steinbuch, K, (1963): Automat und Mensch. Kybernetische Tatsachen und Hypothesen, 3. Aufl., Berlin Heidelberg: Springer

Turing, A.M. (1950): Computing Machinery and Intelligence, Mind 49: 433-460

Tomasello, M. (2009): Die Ursprünge menschlicher Kommunikation, Frankfurt/M: Suhrkamp

Wahlster, W, (2020): Deep Learning alleine reicht nicht, FAZ vom 10.09.2020

Wiener , N. (1948): Cybernetics or control and communication in the animal and the machine, Cambridge (MA): MIT Press

↑Inhalt↑

Die Illusionsfabrik der ›KI‹-Narrative

Derzeit sind medial verbreitete »KI«-Narrative wieder en vogue. In den 1980er Jahren versuchten Ansätze »symbolischer KI«, explizites Wissen über Praktiken kooperativer kognitiver Arbeit und daraus zu ziehende Schlüsse in Gestalt »wissensbasierter« oder »Expertensysteme« zu modellieren. Im Unterschied dazu richten sich heutige Ansätze darauf, zwecks Gewinnung von Berechnungsverfahren zur Bewältigung kognitiver Aufgaben die Mühen analytischer Durchdringung und Modellierung mittels Verfahren »maschinellen Lernens (ML)« zu umgehen – tatsächlich aber nur eine Art Funktions-Approximation an große Mengen vorgegebener Daten. Während erstere an den hohen Hürden hinreichender Analyse und Explizierbarkeit impliziten Wissens gescheitert sind, werfen die neuen Ansätze erneut unüberwindlich erscheinende Probleme auf. Zum besseren Verständnis wird zunächst anhand üblicher »KI«- Definitionen gezeigt, dass ›KI‹-Protagonisten nicht einmal wissen können, worin sich die ›künstlich intelligent‹ genannten Systeme eigentlich genau von herkömmlichen Computersystemen unterscheiden – ein Umstand, aus dem viele Illusionen über Funktionsweisen und Leistungspotenziale dieser Systeme erwachsen. Neue Probleme ergeben sich einerseits aus der kaum einschätzbaren Relevanz und Validität der Daten, zudem aus der Intentionalität und Kontingenz sozialer Praktiken, andererseits aus einer höheren Art der Undurchschaubarkeit des Systemverhaltens. Das wird zudem eine Reihe neuer, freilich noch ungelöster ethischer Fragen auf.

1 Was ist eigentlich ein ›KI‹-System?

Derzeit redet alle Welt nach längerer Pause wieder über »künstliche Intelligenz (KI)« als einer zukunftsweisenden Computertechnik. Angesichts dessen darf angenommen werden, dass einigermaßen klar ist, was diese Technik besonders kennzeichnet, worauf ein Blick auf übliche ›KI‹-Definitionen Auskunft geben sollte. Diese lassen sich in zwei Gruppen einteilen: Die erste Gruppe begreift Computertechnik als ›KI‹- bzw.›AI‹-System, wenn die Lösung der Aufgaben, zu deren Bewältigung es geschaffen wird, natürliche Intelligenz und Erfahrung erfordert:

- »*AI is the part of computer science concerned with ... systems that exhibit characteristics we associate with intelligence in human behaviour – understanding language, learning, reasoning, problem solving, and so on.*« (Barr & Feigenbaum 1981; ähnlich auch schon McCarthy 1955);
- Neuerdings auch: Systems »*that are capable of performing tasks commonly thought to require intelligence. Machine learning ... refers to the develop-*

ment of digital systems that improve their performance on a given task over time through experience.« (Autorengruppe 2018: 9).

Eine zweite Gruppe von Definitionen schreibt ›KI‹-Systemen eine gewisse eigenständige »Handlungsträgerschaft« (»agency«) zu:

- AI research investigates »*intelligent agents*«, i.e. devices »*that perceive their environment and take actions maximizing the chance of successfully achieving their goals*.« (Russell & Norvig 2009: 2);
- »*AI researchers use mostly the notion of rationality, which refers to the ability to choose the best action to take in order to achieve a certain goal, given certain criteria to be optimized and the available resources*.« (High-Level Expert Group on AI 2018: 1 f).

Bei näherem Hinsehen erweisen sich diese Bestimmungen jedoch als reine Scheindefinitionen: Der ersten Gruppe zufolge sollen sich ›KI‹-Systeme von ›gewöhnlichen‹ Computersystemen, selbst anderen technischen Artefakten, dadurch unterscheiden, dass die Bewältigung der Aufgaben, für die sie geschaffen werden, natürliche Intelligenz erfordert. Eben dies gilt aber auch schon für die Lösung relativ einfacher kognitiver Aufgaben wie die Bestimmung der Nullstelle einer quadratischen Gleichung oder das Spiel der »Türme von Hanoi«, die ebenfalls ein beträchtliches Maß an Intelligenz erfordern (das schon die Fähigkeiten vieler Menschen übersteigt, ganz abgesehen davon, dass auch die Ausübung körperlicher Arbeit meist hohe Intelligenz voraussetzt). So wäre diesen Definitionen zufolge auch jedes andere auf einem Computer ausgeführte Berechnungsverfahren ein »KI«- System, die vermeintliche *differentia specifica* unterscheidet nicht wirklich.

Bei der zweiten Definitionsgruppe werden ›KI-Systemen‹ die typischen Merkmale von Intentionalität und rationalem Handeln, die Wahl geeigneter Mittel zum Erreichen von Zielen, einfach zugeschrieben. Tatsächlich befolgen diese aber, wie alle Computersysteme, nur ein ihr Verhalten determinierendes Programm, dem bereits alle denkbaren Bedingungen methodisch eingeschrieben sind, unter denen von außen festgelegte Ziele bestmöglich zu erreichen sind. Hier werden Herstellen und Hergestelltes, die intelligenten Tätigkeiten des Entwerfens und Programmierens mit den Leistungen des Programms als deren Ergebnis verwechselt – ein krasser Kategorienfehler. Tatsächlich vergegenständlicht das Programm lediglich Ergebnisse des lebendigen Arbeitsvermögens, der Intelligenz, Erfahrungen und Fähigkeiten seiner Schöpfer, vorgestellte Ziele unter angenommenen Bedingungen mit Mitteln der Logik und Verfahren der Berechnung bestmöglich zu verwirklichen.

Somit bleibt festzuhalten, dass aufgrund dieser Definitionen niemand wirklich wissen kann, was ein ›KI‹-System eigentlich ist, paradoxerweise auch jene nicht, die ständig davon reden – ein eklatanter Fall von »Technik und Wissenschaft als ›Ideologie‹« (Habermas 1968; vgl. Brödner 1997: Kap. 4.4). Jede Art von Computerprogramm ist schließlich nur eine Vergegenständlichung des

lebendigen Arbeitsvermögens und der Einsichten natürlicher Intelligenz seiner Konstrukteure – eine Feststellung, die freilich auch für jedes andere technische Artefakt gilt, vom Faustkeil bis zum Computer.

2 Die Mühen der Modellierung von Praxis und der vergebliche Versuch ihrer Umgehung

Computersysteme, gleich welcher Komplexität, führen berechenbare Funktionen auf binären Schaltsystemen aus und nichts sonst. Gestaltung und Einsatz erfordern die Modellierung und Formalisierung sozialer Praktiken kooperativer kognitiver Arbeit, ein schwieriger, hohe Einsichtsfähigkeit und Nutzerbeteiligung verlangender Vorgang, der auch leicht misslingen kann (Rohde et al. 2017). Dabei muss die klaffende semantische Lücke zwischen der Praxis und deren sprachlicher Beschreibung einerseits und Programmen als formalen Beschreibungen maschinell ausführbarer Berechnungsverfahren andererseits überwunden werden (Programmiersprachen helfen dabei). Die dafür nötige Modellbildung in aufgabenorientiert reduzierender Perspektive – Kern der Softwaretechnik – durchläuft die folgenden Schritte der Reduktion, Abstraktion und Formalisierung (Andelfinger 1997):

- *Semiotisierung:* Begrifflich-propositionale Beschreibung der Aufgaben und Abläufe einer sozialen Praxis mittels Zeichen liefert ein perspektivisch reduziertes Abbild von Wirklichkeit als Ergebnis gemeinsamer Reflexion und Kommunikation der Akteure (Sprachanalyse, ›Ontologie‹):
  → *Anwendungsmodell*.
- *Formalisierung:* Abstraktion von situations- und kontextgebundenen Bedeutungen und Reduktion auf sinnfreie Standardzeichen und -operationen:
  → *formales Modell* (Spezifikation).
- *Algorithmisierung:* Überführung von Gegenständen und Abläufen des formalen Modells in auto-operational ausführbare Prozeduren in Form von Daten und berechenbaren Funktionen (Algorithmen):
  → *Berechnungsmodell* (als Grundlage der Programmierung)

Sprachlich repräsentierte Vorgänge kooperativer kognitiver Arbeit können so partiell formalisiert und dann als berechenbare Funktionen (Algorithmen) maschinell ausgeführt werden – auch Menschen rechnen formalisiert wie Maschinen, ihre Fähigkeiten sind aber nicht darauf beschränkt (daher gilt der Einsatz von Computern auch als »Maschinisierung von Kopfarbeit«; Nake 1992).

Die Ausführung der berechenbaren Funktionen stellt einen ›degenerierten‹, auf eine dyadische Relation reduzierten Zeichenprozess ohne ›Fenster zur Welt‹ dar, dem der Bezug zu einem erlebten, leiblich erfahrenen oder gedachten Objekt, eben die ›Be-zeichnung‹ fehlt. Es ist nur eine »Quasi-Semiose«, die mit Signalen (logisch: ›Daten‹) als auf Syntax reduzierten »Quasi-Zeichen« operiert (Nöth 2002). Deren Zustände werden per Programm rein physisch transformiert ohne Ansehen von Bedeutung. Im Computersystem implementiert entstehen damit

»autooperationale Formen« (Floyd 2002) als Ausdruck abstrakter, formalisierter Operationen. Deren Sinn muss durch Aneignung seitens der Systemnutzer für wirksamen praktischen Gebrauch erst noch erschlossen werden.

Dabei ist zwischen Problem und Aufgabe zu unterscheiden (Dörner 1983): Ein Problem liegt vor, wenn die Mittel zum Erreichen eines angestrebten Ziels noch unbekannt sind oder über das Ziel keine klaren Vorstellungen bestehen, wenn handelnde Personen also nicht wissen, wie sie ihr Ziel erreichen sollen: »Intelligenz ist das, was man einsetzt, wenn man nicht weiß, was man tun soll.« (J. Piaget). Gesucht sind dann Ideen für abduktives Schließen, d.h. die Bildung von erklärenden Hypothesen aufgrund von Intuition, Analogie oder Kreativität (Peirce 1878). Bewährt sich eine Hypothese, können damit Verfahren zur methodischen Bewältigung der dem Problem entsprechenden Aufgaben gewonnen werden (Popper 1994).

Davon unterscheiden sich Aufgaben als geistige Anforderungen, für deren Bewältigung Methoden oder Verfahren bereits existieren. Aufgaben erfordern lediglich den Einsatz bekannter Mittel auf gewohnte Weise; als Instanzen eines prinzipiell bereits gelösten Problems erfordert ihre Lösung lediglich den routinierten Gebrauch dafür angeeigneter Methoden oder Verfahren (einschließlich der Beurteilung ihrer jeweiligen Eignung).

Die Modellierung einer komplexen sozialen Praxis beginnt als Problemlösung: Anfangs sind weder das Problem noch dessen Lösung hinreichend durchschaut; sie müssen im Zuge der Semiotisierung erst durch Analyse und Genese expliziten Wissens verstanden werden, um gesicherte Methoden der Bewältigung zu gewinnen. Dadurch wird die weitere Modellierung zur Aufgabe reduziert und durch Anwendung des Lösungsverfahrens bewältigt. In der Problemanalyse, der Wissensgenese, der Schaffung formalisierter Lösungsverfahren und der Beurteilung ihrer Eignung erweist sich die natürliche Intelligenz der Akteure, während die Leistung des Computersystems auf die Ausführung des daraus entstandenen programmierten Berechnungsmodells beschränkt ist, ggf. unter Berücksichtigung äußerer Bedingungen.

Mit der derzeit im Zentrum des Interesses stehenden Verfahren »maschinellen Lernens« und der Nutzung von »Big Data« wird versucht, sich diese Mühen von Problemanalyse, Modellierung, Formalisierung und Bestimmung eines spezifischen Berechnungsmodells zu ersparen. Stattdessen werden einfach für ganze Klassen von Aufgaben – darunter Aufgaben der Objekt-Klassifizierung, der Clusterung von Objekten oder automatisierter Entscheidung – erfahrungsbasiert oder schlicht aufgrund theorielosen Probierens geeignet erscheinende, generische mathematische Funktionen ausgewählt, deren offene Parameter noch aufgabenspezifisch zu bestimmen sind. Solche Funktionen sind etwa ›künstliche neuronale Netze (KNN)‹ mit ihren Gewichten, Polynome bzw. logistische Funktionen mit ihren Koeffizienten oder Entscheidungsbäume mit ihren Kantengewichten als Parametern.

Die Parameter werden mittels meist längst bekannter Verfahren der Funktions-Approximation möglichst gut an große Mengen verfügbarer Datenobjekte angepasst, was sie als ›adaptive Systeme‹ kennzeichnet. Die so für die Bewältigung einer spezifischen Aufgabe »trainierten« Funktionen lassen sich auf neue Datenobjekte gleicher Art anwenden, vorausgesetzt, der infolge prinzipieller Kontingenz sozialer Praktiken veränderliche Kontext bleibt erhalten. Dieses Vorgehen mag in je besonderen Einzelfällen durchaus gelingen, setzt aber meist enorme Rechenleistung voraus (die jüngst erst verfügbar ist). Diese Art »maschinellen Lernens« hat aber nichts mit herkömmlichem Verständnis reflexiven, auf Einsicht beruhenden Lernens zu tun und ist insofern eine irreführende Benennung. Der Erfolg steht und fällt mit den zum ›Training‹ benutzten Daten, deren Herkunft und Qualität aber meist nicht einschätzbar und hinsichtlich Repräsentativität und Verzerrungen (Biases) oft äußerst fragwürdig sind.

Ein solches Vorgehen hat den hohen Preis, dass grundsätzlich nur wahrscheinliche, von der Vorgeschichte abhängige, daher stets unsichere Ergebnisse zu erwarten sind, deren Validität kaum zu beurteilen ist – eine Art postmoderner Obskurantismus, unreflektierter Datengläubigkeit geschuldet. Den berechneten Ergebnissen kann man nur blind vertrauen, weil sich aktual nicht nachvollziehen lässt, wie sie im einzelnen zustande gekommen und wie zuverlässig sie sind. Solch neue Art undurchschaubaren Systemverhaltens hat im Gebrauch allerdings höchst abträgliche Folgen, die reflexives Lernen behindern und großes Stresspotenzial aufweisen.

3 Ungelöste ethische Fragen

Die hinsichtlich Validität, Repräsentativität und Aktualität oftmals unsichere Qualität der Daten über zugrunde liegende reale Vorgänge wirft, zusammen mit der Intransparenz und Variabilität des Systemverhaltens und der prinzipiellen Unsicherheit berechneter Ergebnisse, schwerwiegende ethische Fragen auf (Mittelstadt et al. 2016): Dürfen derartige potenziell gefährliche Systeme überhaupt in praktischen Einsatz gelangen? Wie lassen sich deren Sicherheit und (auch nicht intendierten) Schadenspotenziale im Vorhinein bewerten und wer wird im Schadensfall zu auch haftender Verantwortung gezogen – die Hersteller, die Betreiber oder gar einzelne Nutzer? Dazu werden etwa unter dem Stichwort »trustworthy AI« zwar weithin allgemeine Bewertungskriterien diskutiert, fraglich bleibt jedoch, ob wegen genannter Systemeigenschaften konkrete Regelungen überhaupt verbindlich festlegbar sind (vgl. z.B. High-Level Expert Group on AI 2019) – man fragt sich, ob nicht hinter der Fassade großer ethischer Besorgnis riskante Entwicklungen einfach weiter betrieben werden sollen.

Zudem ist, in Anbetracht der begrenzt erscheinenden Möglichkeiten vollständiger Automatisierung ganzer Arbeitsprozesse, auf absehbare Zeit mit der Notwendigkeit des Zusammenwirkens von Wissensarbeitern mit adaptiven Syste-

men zu rechnen. Wegen deren undurchschaubaren Eigenlebens ist statt bislang üblicher Interaktion aber nur noch Ko-Aktion möglich, die den zweckorientierten instrumentellen Gebrauch der Systeme erheblich erschwert und deren Nutzer mit beträchtlichen Handlungshindernissen konfrontiert: Unter dem Druck zugewiesener Leistungsforderungen und Eigenverantwortung können sie den Resultaten mangels Urteilsfähigkeit nur blind vertrauen, leiden mithin unter großer Unsicherheit, jedoch ohne die Möglichkeit, sich das Systemverhalten hinreichend aneignen zu können (vgl. den Beitrag »Paradoxien der Ko-Aktion…«).

Das führt gesicherten arbeitswissenschaftlichen Erkenntnissen zufolge zu beträchtlichen spezifischen Belastungen und Stressreaktionen (Brödner 2020). Dementsprechend wird auch von vielen Seiten vehement gefordert, die Systeme mit Komponenten auszurüsten, die das Zustandekommen ihrer Resultate auf Verlangen mit hinreichender Detailwiedergabe zu erklären und damit auch den instrumentellen Gebrauch zu erleichtern vermögen (»explainable AI«). Deren Realisierung steht aber, so überhaupt möglich, noch in weiter Ferne und solange es sie noch nicht gibt, sollte der Einsatz adaptiver Systeme im Interesse effizient und sozialvertraglich gestalteter Arbeitsprozesse unbedingt vermieden werden.

4 Fazit

Während derzeit viel und durchaus zurecht von ethischen Herausforderungen durch »ML«-Systeme die Rede ist, scheinen tragfähige Lösungen noch in weiter Ferne zu liegen. Insbesondere lassen praxistaugliche Ergebnisse der Bemühungen um eine »explainable AI« noch auf sich warten, die der Intransparenz geschuldete unzumutbare Stresssituationen für die Nutzer zu vermeiden in der Lage wären. Dessen ungeachtet verleiten gängige, trotz entgegenstehender Erkenntnisse ständig reproduzierte ›KI‹-Erzählungen über vermeintlich »lernfähige« oder gar »autonome« Systeme zu folgenreichen Illusionen über deren tatsächliche Leistungsfähigkeit. Sie sind weder »lernfähig«, passen sich allenfalls mittels gegebener Daten algorithmisch gesteuert an äußere Gegebenheiten an, noch »autonom«, d.h. in der Lage, eigene Funktionsregeln zu setzen, sondern sind, wie jedes andere selbsttätige Computersystem auch, per Programm fremdgesteuerte Automaten (oft raffiniert ausgedacht, wie Hofstadter bereits (1979: 601) spottete: »AI is whatever hasn't been done yet«). Damit erweisen sich ›KI‹-Erzählungen als in der Sache unbegründete, durch falsche Begriffsbildung geschaffene Brutstätten gefährlicher und Ressourcen fehlleitender Illusionen (was naiven Rezipienten freilich entgeht).

Für einige in das Geschehen involvierte Akteure sind diese Illusionen jedoch durchaus verlockend: Politiker können sich damit als Förderer von »Modernisierung« profilieren, Forschungsstätten erhalten reichlich Mittel, die dem Nachwuchs viele Promotionsthemen bieten und Unternehmen vermögen sich vorübergehend neue Geschäftsfelder zu erschließen. Das Ergebnis ist allerdings Science Fiction – wirkmächtige Fiktion und miserable ›Science‹. Und die ange-

sprochenen Probleme lassen eher unerwartet großen Aufwand bei minimalem Ertrag wenn nicht gar das Menetekel eines erneuten Scheiterns erahnen.

5 Literatur

Andelfinger, U. (1997): Diskursive Anforderungsanalyse. Ein Beitrag zum Reduktionsproblem bei Systementwicklungen in der Informatik, Frankfurt/M: Peter Lang

Autorengruppe (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, Oxford (AR): Future of Humanity Institute u.a. 02/2018, hhps://arxiv.org/pdf/1802.07228.pdf

Barr, A. & Feigenbaum, E.A. (1981): The Handbook of Artificial Intelligence, Stanford (CA): HeurisTech Press

Brödner, P. (2020): Paradoxien der Koaktion von Experten und adaptiven Systemen, in: P. Brödner & K. Fuchs-Kihowski (Hg.): Zukunft der Arbeit – soziotechnische Gestaltung der Arbeitswelt im Zeichen von »Digitalisierung« und »Künstlicher Intelligenz«, Abhandlungen der Leibniz-Sozietät der Wissenschaften Band 67, Berlin: trafo Wissenschaftsverlag, 143-159

Brödner, P. (1997): Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen, Berlin: edition sigma

Dörner, D. (1983): Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität. Bern: Huber

Floyd, C. (2002): Developing and Embedding Autooperational Form, in: Dihrich, Y.; Floyd, C.; Klischewski, R. (Eds.): Social Thinking – Software Practice, Cambridge (MA): MIT Press, 5-28

Habermas, J. (1968): Technik und Wissenschaft als ›Ideologie‹, Frankfurt/M: Suhrkamp

Haug, W. F. (2005): Vorlesungen zur Einführung ins »Kapital«, Hamburg: Argument

High-Level Expert Group on Artificial Intelligence (2018): A Definition of AI: Main Capabilities and Scientific Disciplines, Brussels: European Commission

High-Level Expert Group on Artificial Intelligence (2019): Ethics Guidelines for Trustworthy AI, Brussels: European Commission

Hofstadter, D.R. (1979): Gödel, Escher, Bach. An Eternel Golden Braid, New York: Vintage Books

McCarthy, J. (1955): A Proposal for the Summer Research Project on Artificial Intelligence, hhp://www-formal.stanford.edu/jmc/history/dartmouth.pdf

Mittelstadt, B. D.; Allo, P.; Taddeo, M.; Wachter, S. & Floridi, L. (2016): The Ethics of Algorithms: Mapping the Debate, Big Data & Society 3 (2), 1-21

Nake, F. (1992): Informatik und die Maschinisierung von Kopfarbeit, in: Wolfgang Coy et al. (Hg.): Sichtweisen der Informatik, Braunschweig Wiesbaden: Vieweg, 181-201

Nöth, W. (2002): Semiotic Machines, Cybernetics and Human Knowing 9 (1), 5-22

Peirce, C.S. (1878): Deduction, Induction, and Hypothesis, in: Collected Papers Vol. 2, ed. by C. Hartshorne, & P. Weiss, P., Cambridge (MA): Harvard University Press (1931-35)

Popper, K.R. (1994): Alles Leben ist Problemlösen. Über Erkenntnis, Geschichte und Politik, München. Piper

Rohde, M.; Brödner, P.; Stevens, G.; Betz, M. & Wulf, V. (2017): Grounded Design – a Praxeological IS Research Perspective, Journal of Information Technology 32 (2), 163-179

↑Inhalt↑

Publikationsnachweise

Coping with Descartes' Error in Information Systems, ursprünglich erschienen in: AI & Society Journal of Knowledge, Culture and Communication, published online January 17, 2018

›Super-intelligent‹ Machine: Technological Exuberance or the Road to Subjection, ursprünglich erschienen in: AI & Society Journal of Knowledge, Culture and Communication, published online May 29, 2017

Industrie 4.0 und Big Data – wirklich ein neuer Technologieschub? Ursprünglich erschienen in: H. Hirsch-Kreinsen; P. Ittermann & J. Niehaus (Hg.): Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen, 2. überarbeitete Aufl., Baden-Baden: Nomos 2018, 323-346

Grenzen und Widersprüche der Entwicklung und Anwendung ›Autonomer Systeme‹, ursprünglich erschienen in: H. Hirsch-Kreinsen & A. Karačić (Hg.): Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt, Bielefeld: transcript 2019, 69-97

›Machines that think‹ – die »KI«-Illusion und ihre Wurzeln, ursprünglich erschienen in: Klaus Lenk & Jörg Pohle (Hg.): Der Weg in die »Digitalisierung« der Gesellschaft. Was können wir aus der Geschichte der Informatik lernen? Marburg: Metropolis 2021, 67-82

Paradoxien der Ko-Aktion von Experten und adaptiven Systemen, ursprünglich erschienen in: P. Brödner & K. Fuchs-Kittowski (Hg.): Zukunft der Arbeit – soziotechnische Gestaltung der Arbeitswelt im Zeichen von »Digitalisierung« und »Künstlicher Intelligenz«, Abhandlungen der Leibniz-Sozietät der Wissenschaften Band 67, Berlin: trafo Wissenschaftsverlag 2020, 143-159

Das Produktivitätsparadoxon der Computertechnik, ursprünglich erschienen in: H. J. Bontrup & J. Daub (Hg.): Digitalisierung und Technik – Fortschritt oder Fluch? Perspektiven der Produktivkraftentwicklung im modernen Kapitalismus, Köln, PapyRossa 2020, 114-144

›Informatik‹ – eine Wissenschaft auf Abwegen, ursprünglich erschienen in:
G. Banse & K. Fuchs-Kittowski (Hg.): Cyberscience – Wissenschaftsforschung und Informatik. Digitale Medien und die Zukunft der Kultur wissenschaftlicher Tätigkeit, Sitzungsberichte der Leibniz-Sozietät, Band 150/151, Berlin: trafo Wissenschaftsverlag 2022, 257-273

Die Illusionsfabrik der ›KI‹-Narrative, ursprünglich erschienen in: FIfF Kommunikation 39 (2), 2022, 32-36