Plappern, sondern auch teilweise zu philosophisch sinnvollen Antworten bringen konnte. Allerdings, so lässt sich einwenden, weiß GPT-3 nichts davon, ob es sinnvolle oder unsinnige Antworten gibt, und Menschen sollten die für sie sinnvollen nicht als systemimmanent sinnvoll missdeuten.

Vladova und Friesike konzentrieren sich auf das Lernen aus Irrtümern als wichtigen, für KI unerreichbaren Aspekt menschlicher Intelligenz. Auch dagegen lässt sich einwenden, dass die Methode beim Backtracking im Deep Learning gerade das Annähern an fehlerhafte Ergebnisse ist. Aber es ist wohl gemeint, dass Fehler neue oder unerwartete Lösungen in Gang setzen.

Sybille Krämer lenkt den Blick auf einfache Anwendungen von schwacher KI (wobei sie diese mit allgemeiner Software identifiziert) und sieht für diese eine weitere Verflachung der menschlichen Kultur in Text und Bild auf das Maschinenlesbare. Im Weiteren problematisiert sie innerhalb der KI das Maschinelle Lernen und stellt sehr in Frage, ob es je möglich wäre, dass Maschinen Sinn und Bedeutungszusammenhänge verstehen können sollten.

Rico Hauswald vergleicht Deep-
keln und sieht darin Vorteile (Dem
Gefahren, da die Black Box von U
dings ist beides nicht KI-typisch, d
klassische Entscheidungsalgorithmen. Thomas Weiss hinwiederum diskutiert die Vollautomatisierung im Hinblick auf Arbeitsplatzverluste und phantasiert über sich emanzipierende Artificial General Intelligence von einem KI-Proletariat.

Reinhard Kahle problematisiert die rückwärtsgewandten Entscheidungen durch KI, wenn sie auf vergangenen Datensammlungen beruhen, beispielsweise im Hinblick auf ein Ausbleiben von Innovationen sowie den Autonomieverlust von Menschen bei der Übernahme von Aufgaben durch KI.

Engel und Schultheis befassen sich ebenfalls mit KI-basierten Entscheidungen anhand einer Delphi-Studie zu den Erwartungen der Befragten über die entsprechenden Veränderungen. Daran anschließend erörtern sie, wieviel Vertrauen in diese Technologie gesetzt werden kann oder sollte, was sie weiter zu moralischen Fragestellungen führt.

Mit Ethik und Moral im Zusammenhang mit KI befassen sich die weiteren Texte, wobei zu unterscheiden ist, ob sich die ethischen Fragen auf KI-Ergebnisse beziehen, oder ob versucht werden soll, ethische Anforderungen ins Design von KI-Entwicklungen einzubeziehen. *John Michael und Hendrik Kempt* diskutieren den Aufbau von Beziehungen und Vertrauen zwischen Menschen und Robotern in Grenzen durchaus positiv, während *Orphelia Deroy* gerade die Möglichkeit von Vertrauen in KI problematisiert und dazu Rechtfertigung verlangt. *Catrin Misselhorn* hält hingegen eine Maschinenethik für möglich, fordert dazu aber drei Grundsätze, von denen zwei meines Erachtens im Widerspruch zur Funktionsweise von KI stehen und so ihre Anwendung aushebeln würden, etwa wenn der Mensch immer die Verantwortung für KI-Entscheidungen tragen soll.

Ein wenig enttäuschend ist, dass das Buch wenig auf die Funktionsweise von KI zum Verständnis ihrer Möglichkeiten und Unmöglichkeiten eingeht. Ausnahmen bilden nur Hans-Jörg Kreowskis und Wolfgang Kriegers Darstellung der Entscheidbarkeit und Berechenbarkeit S. 272 ff. und Sybille Krämers Darstellung von Deep Learning ab S. 346, dies alles auch nur im Bereich des
mir erstaunlich, dass die Körper-
e von Computern und Serverfar-
en ungeheuren Konsum von und
r und seine Erzeugung, Metallab-
bau und seine Isolation und Bearbeitung nicht in Betracht gezogen wird. Ganz besonders fehlen solche Überlegungen bei jenen KI-Phantasten, die die Ablösung des Menschen durch eine höhere Intelligenz erwarten oder befürchten, als ob KI-Systeme selbständig in der Erde nach seltenen Erden buddeln, Chips herstellen oder Windräder zur Stromerzeugung aufstellen würden. Vielleicht sollen das alles dann Menschensklaven in einer KI-Diktatur erledigen. Aber mit welchen Druckmitteln könnte KI die Menschen dazu knechten?

erschienen in der FIfF-Kommunikation, herausgegeben von FIfF e.V. - ISSN 0938-3476 www.fiff.de

Anmerkung

1 *https://www.sueddeutsche.de/kultur/clemens-j-setz-georg-buechner-preis-woyzeck-karl-krall-denkende-tiere-pferde-1.5457889*

Peter Brödner

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 wirft zudem eine Reihe neuer, freilich noch ungelöster ethischer Fragen auf.*

Was ist eigentlich ein ›KI‹-System?

Derzeit redet alle Welt nach längerer Pause wieder über *künstliche Intelligenz (KI)* als zukunftsweisende Computertechnik. Angesichts dessen darf angenommen werden, dass einigermaßen klar ist, was diese Technik besonders kennzeichnet, worüber 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 development 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.

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 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 nötige Modellbildung in aufgabenorientiert reduzierender Perspektive – Kern der Softwaretechnik – durchläuft folgende 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 *auto-operationale 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 oder 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 undurchschaubares Systemverhalten hat im Gebrauch allerdings höchst abträgliche Folgen, die reflexives Lernen behindern und großes Stresspotenzial aufweisen.

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 damit zu rechnen, dass Wissensarbeiter mit adaptiven Systemen zusammen werden arbeiten müssen. Wegen deren undurchschaubaren Eigenlebens ist statt bislang üblicher Interaktion aber nur noch Ko-Aktion möglich, die 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. 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 sozialverträglich gestalteter Arbeitsprozesse unbedingt vermieden werden.

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*, also 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 angesprochenen Probleme lassen eher unerwartet großen Aufwand bei minimalem Ertrag, wenn nicht gar das Menetekel eines erneuten Scheiterns erahnen.

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, https://arxiv.org/pdf/ 1802.07228.pdf

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

Brödner P (2020) Paradoxien der Koaktion von Experten und adaptiven Systemen, in: Brödner P, Fuchs-Kittowski K Hg.: Zukunft der Arbeit – sozio-technische 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: Dittrich 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 WF (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

Peter Brödner

Peter Brödner, Prof. Dr.-Ing., Jahrgang 1942, Studium des Maschinenbaus in Karlsruhe und Berlin. Nach verschiedenen Stationen in Forschung und Projektmanagement auf dem Gebiet computerunterstützter Produktion bis 2005 Forschungsdirektor für Produktionssysteme am Institut Arbeit und Technik im Wissenschaftszentrum Nordrhein-Westfalen. Seither im Ruhestand, Honorarprofessor an der Universität Siegen (Wirtschaftsinformatik), Mitglied der Leibniz-Sozietät der Wissenschaften zu Berlin.

High-Level Expert Group on Artificial Intelligence (2019) Ethics Guidelines for Trustworthy AI, Brussels: European Commission

Hofstadter DR (1979) Gödel, Escher, Bach. An Eternal Golden Braid, New York: Vintage Books

McCarthy J (1955) A Proposal for the Summer Research Project on Artificial Intelligence, http://www-formal.stanford.edu/jmc/history/dartmouth.pdf

Mittelstadt BD, 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 CS (1878) Deduction, Induction, and Hypothesis, in: Collected Papers, Vol. 2, ed. by Hartshorne C, Weiss P, Cambridge (MA): Harvard University Press (1931-35)

Popper KR (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

Rainer Rehak

The Language Labyrinth: Constructive Critique on the Terminology Used in the AI Discourse

Introduction

In the seventies of the last century, the British physicist and science fiction writer Arthur C. Clarke coined the phrase of any sufficiently advanced technology being indistinguishable from magic – understood here as mystical forces not accessible to reason or science. In his stories Clarke often described technical artefacts such as anti-gravity engines, 'flowing' roads or tiny atom-constructing machinery. In some of his stories, nobody knows exactly how those technical objects work or how they have been constructed, they just use them and are happy doing so.

In today's specialised society with a division of labour, most people also do not understand most of the technology they use. However, this is not a serious problem, since for each technology there are specialists who understand, analyse and improve the products in their field of work – unlike in Clarke's worlds. But since they are experts in few areas and human lifetime is limited, they are, of course, laypersons or maybe hobbyists in all other areas of technology.

After the first operational universal programmable digital computer – the Z3 – had been invented and built in 1941 in Berlin by Konrad Zuse, the rise of the digital computer towards today's omnipresence started. In the 1960s, banks, insurances and large administrations began to use computers, police and intelligence agencies followed in the 1970s. Personal computers appeared and around that time newspapers wrote about the upcoming 'electronic revolution' in publishing. In the 1980s professional text work started to become digital and in the 1990s the internet was opened to the general public and to commercialisation. The phone system became digital, mobile internet became available and in the mid-2000s smartphones started to spread across the globe (Passig and Scholz 2015).

During the advent of computers, they were solely operated by experts and used for specialised tasks such as batch calculations and book-keeping at large scale. Becoming smaller, cheaper, easier to use and more powerful over time, more and more use cases emerged up to the present situation of computer ubiquity. More applications, however, also meant more impact on personal lives, commercial activities and even societal change (Coy 1992). The broader and deeper the effects of widespread use of networked digital computers became, the more pressing political decisions about their development and regulation became as well.

The situation today is characterised by non-experts constantly using computers, sometimes not even aware of it, and non-experts making decisions about computer use in business, society and politics – from schools to solar power, from cryptography to cars. The only way to discuss highly complex computer systems and their implications is by analogies, simplifications and metaphors. However, condensing complex topics into understandable, discussable and then decidable bits is difficult in at least two ways. First, one has to deeply understand the subject and second, one has to understand its role and context in the discussion to focus on the relevant aspects. The first difficulty is to do with knowledge and lies in the classical technical expertise of specialists. But the second difficulty concerns what exactly should be explained in what way. Depending on the context of the discussion, certain aspects of the matter have to be explicated using explanations, metaphors and analogies highlighting the relevant technical characteristics and implications. Seen in this light, this problem of metaphors for technology is not only philosophically highly interesting but also politically very relevant. Information technology systems are not used because of their actual technical properties, but because of their assumed functionality, whereas the discussion about the functionality is usually part of the political discourse itself (Morozov 2013).

Given the complexity of current technology, only experts can understand such systems, yet only a small number of them actively and publicly take part in corrective political exchanges about technology. Especially in the field of artificial intelligence (AI) a wild jungle of problematic terms is in use. However, as long as discussions take place among AI specialists those terms function just as domain-specific technical vocabulary and no harm is done. But domain-specific language often diffuses into other fields and then easily loses its context, its specificity and its limitations. In this process terms which might have started as