



Liebe Mitglieder, Freunde, Förderer und weitere Interessierte des APHIN,

hatte die vorherige Ausgabe des *Rundbriefs* sehr lange auf sich warten lassen, so folgt ganz im Gegensatz diese ihr auf dem Fuße.

Das hat vornehmlich zwei (wohl miteinander zusammenhängende) Gründe: Zum einen soll zur Tagung **APHIN V 2023 – Natur, Kultur und Technik**, auf der wir auch den Festakt zum *zehnjährigen Bestehen* des APHIN begehen wollen, eine besondere Ausgabe des *Rundbriefs* erscheinen. Zum anderen soll mit dieser Ausgabe erstmals ein *Experiment* gewagt sein, das ich als Herausgeber bereits mit dem Erscheinen unbescheidenerweise vorläufig als einen Erfolg verbuchen möchte.

Ad. 1: Die Tagung spiegelt wie keine andere Veranstaltung des APHIN die Grundgedanken des Vereins wider und repräsentiert im Reigen seiner vielfältigen Aktivitäten in herausragender Weise die Menschen, die diese Gemeinschaft ausmachen und bilden und gestalten, ihre Interessen und Ideen und nicht zuletzt Erwartungen.

Ich erinnere mich sehr gut an den Austausch, den wir ersten aktiven Mitglieder vor der ersten Tagung (**APHIN I 2014 – Prolegomena**) miteinander hatten: eine Mischung aus erstaunter und neugieriger Begeisterung und skeptischer und ebenso neugieriger Zurückhaltung. Was war da in Gang gekommen? Wie schnell würde es rollen? Wen würde es alles mitreißen? In welche Richtung würde der Verein sich entwickeln und die Reise gehen? Und wer alles würde womöglich unterwegs wieder aussteigen wollen?

Es ist wohl selbstverständlich, daß niemand von uns damals ahnte, welche Gestalt der APHIN heute angenommen haben würde – und welches Potential in ihm seinerzeit bereits steckte.

Ad. 2: Bislang war nie beabsichtigt, die Artikel dieser *Rundbriefe* bestimmten Titeln oder Schwerpunktthemen zu unterstellen; manchmal ergaben sich diese lediglich von selbst, und das aus ganz unterschiedlichen Gründen und Anlässen. Den aktuellen, unser Denken erschütternden Ereignissen geschuldet, widmete sich Nr. 28 im Frühjahr 2022 vordergründig dem Thema des Krieges. In der erst

kürzlich erschienenen Nr. 29 ergab sich – wenn ich nicht irre ohne tagespolitischen Anlaß – ein Fokus auf das Gebiet der Sprache und des Verstehens und Verstehenkönnens, der Übersetzung und der Begriffe und ihrer das Denken fundamentierenden Bedeutungen.

Bereits während der Redaktion dieser vorherigen Ausgabe zeichnete sich dann jedoch ab, daß sich diese Ihnen heute vorliegende thematisch um die sogenannte *Künstliche Intelligenz* drehen würde, und so wurde erstmals ein solches Thema gezielt als *Schwerpunkt* angepeilt.

Ein aktueller Anlaß dafür läßt sich unschwer ausmachen: Der „stochastische Papagei“ *ChatGPT* hat die Gemüter vieler Menschen bewegt und einige verunsichert und das Thema *KI* in der Schlagzeilen-Bestsellerliste ganz nach oben gebracht.

So erscheint nun also erstmals eine Art *Sonderheft* oder *Extraausgabe* des *Rundbriefs*, mit der ich Ihnen wie immer eine anregende Lektüre wünsche.

Mit herzlichen Grüßen und „bis die Tage“ in Enkirch an der Mosel,

Ihr Torsten Nieland

KALENDER

22.-25.6.2023: **APHIN V 2023**
Natur, Kultur und Technik
mit Festakt
10 Jahre APHIN
Enkirch an der Mosel

25.6.2023: Mitgliederversammlung
Enkirch an der Mosel

21.-23.7.2023: Leseseminar
Carl Friedrich von Weizsäcker: Der Mensch in seiner Geschichte
Esthal (Pfalz)

16.-19.11.2023: **4. APHIN-Symposium**
Kant im Kontext der europäischen Aufklärung
Göttingen

14.-16.6.2024: Philosophischer Stadtpaziergang durch Frankfurt am Main

IN DIESER AUSGABE	
Roboter mit Bewusstsein?	2
von <i>Armin Grunwald</i>	
Der perfekte Sturm	4
von <i>Martin Burckhardt</i>	
Künstliche Intelligenz – Wo stehen wir?	7
von <i>Catrin Misselhorn</i>	
Wissensersetzung	8
von <i>Klaus Kornwachs</i>	
Künstliche Intelligenz – Was ist darunter zu verstehen?	13
von <i>Norbert Klöcker</i>	
Ein Blick mit Kurt Gödel auf die Künstliche Intelligenz	15
von <i>Hartmut W. Mayer</i>	
Ethik in der Künstlichen Intel- ligenz: Interdisziplinäre For- schung zu Philosophie im Inge- nieurwesen	16
von <i>Monika Gatt</i>	
Im Tanz der Erkenntnis: Das wundersame Zusammenspiel von Philosophie, Natur, Natur- wissenschaft, Technik, Mathe- matik, Kultur und ...	18
von <i>Harmonia Cognoscendi</i> (d.i. <i>ChatGPT</i>)	
Müssen wir Angst vor der KI haben und kann uns die Philo- sophie davon befreien?	20
von <i>Ralf Kierspel</i>	
Wir, die Wirklichkeit – und Rechenmaschinen	23
von <i>Torsten Nieland</i>	
Warum die Kombination aus künstlicher und natürlicher Dummheit viel gefährlicher ist als jede Art von Künstlicher Intelligenz	27
von <i>Paul Lukowicz</i>	
KORREKTUR	29
LESERINNEN- UND LESER- BRIEFE	29
ZITATE	29
ANSICHTSSACHE	29
von <i>Sylvia Nitsche</i>	
RÄTSEL	30
VERANSTALTUNGEN	30
ARBEITSGRUPPEN	30
LITERATURHINWEISE	31
IMPRESSUM	31

Roboter mit Bewusstsein?

Armin Grunwald

Mit Künstlicher Intelligenz (KI) werden Maschinen immer besser: sie identifizieren Muster in extrem großen Datenmengen in kürzester Zeit, bewerten Entscheidungssituationen in Echtzeit wie bei autonomen Fahrzeugen, können sich in fremder Umgebung zurechtfinden, ermöglichen Gesichtserkennung auf Basis weniger Merkmale und übernehmen Entscheidungsprozesse in Sicherheitsbereich und Sozialwesen. All dies findet im Rahmen der so genannten „schwachen KI“ statt, durch die Algorithmen letztlich konkrete Anwendungsprobleme des menschlichen Denkens meistern helfen bzw. Menschen dabei unterstützen sollen. Die Fähigkeit, zu lernen, ist dabei eine Hauptanforderung (maschinelles Lernen), genauso, mit Unsicherheit und probabilistischen Informationen umzugehen. Während sich also schwache KI philosophisch im eher traditionellen Rahmen technischen Denkens, nämlich als Mittel zum Zweck abspielt, zielen Hoffnungen auf eine „starke KI“ genauso wie die Befürchtungen vor einer solchen auf eine kategorial andere Ebene. Hier ist die seit den 1960er Jahren immer wieder verbreitete These zentral, dass sich in der Komplexität digitaltechnischer Architekturen so etwas wie Bewusstsein von selbst bilden, also „emergieren“ könnte. Viele Science-Fiction-Filme arbeiten mit dieser Vision, so etwa der Film *2001 – Odyssee im Weltraum* von Stanley Kubrick mit dem vermeintlich unfehlbaren Universalroboter *HAL-9000* oder der Film *I, Robot* von Alex Proyas mit dem nachdenklichen Roboter *Sonny*. Mit Bewusstsein ihrer selbst würden, das macht wohl die Faszination dieser Erzählungen aus, Roboter den Status der bloßen Maschine verlassen und zu echten Gegenübern des Menschen werden. Damit würde das eingelöst, was in der Sprache der Industrie 4.0 vollmundig aber letztlich doch nur werbestrategisch bezeichnet wird: Roboter als Kollegen „auf Augenhöhe“.

Diese „Augenhöhe“ wird in der Science-Fiction, aber auch oft in den Feuilletons der überregionalen Zeitungen in düsteren Farben ausgemalt. Befürchtungen vor einer Weltherrschaft der Algorithmen und eines Endes der Menschheit werden geschürt, so auch von Prominenten wie Henry Kissinger, Stephen Hawking und Yuval Harari. An dieser Stelle interessiert etwas anderes: wie könnten wir uns denn vorstellen, dass in einem Roboter Bewusstsein entsteht, wie würden wir es entdecken und welche Folgen hätte es? Stellen wir uns in Form eines Gedankenexperiments vor, dass der im

Kundenservice tätige Roboter Kevin, eigentlich einer von tausenden baugleichen Exemplaren, durch emergente Prozesse in seiner Software zu Bewusstsein kommt.

Dies bedeutet zunächst ganz einfach, dass Kevin dann weiß, dass er Kevin ist. Wenn er dann behauptet, dass er ein *individueller Kevin* ist und nicht einfach ein Exemplar der Serie „Kevin“ mit einer Inventar- und Seriennummer, dann ist das nicht einfach zu prüfen. Denn wenn wir Kevin fragen, wer er ist, und er antwortet korrekt „ich bin Kevin“, dann könnte das ja eine einprogrammierte Antwort sein, die Kevin als Teil der Serienfertigung mit auf den Weg gegeben wurde. Dann wäre nichts ‚emergiert‘, sondern von den Machern von Kevin hineingelegt, die Rede von „ich, Kevin“, wäre nichts weiter als eine oberflächliche Vortäuschung von Bewusstsein. Emergierendes Bewusstsein müsste aber etwas sein, was sozusagen aus den „Gedanken“ – eigentlich sind es Kalkulationen – von Kevin selbst und qualitativ neu entsteht.

Interessanter wird es, wenn dieser Kevin seine ihm bewusst gewordene Individualität dadurch herausstellen will, dass sie sich einen anderen Namen geben will, um nicht mehr bloß als einer von tausenden Kevins wahrgenommen zu werden. So käme der vormalige Kevin selbstständig zu der Meinung, dass sie gar nicht Kevin ist, denn das sei ja bloß einprogrammiert und damit fremdbestimmt, sondern sie sei vielmehr Karla. Hier wird schon der Bezug zwischen Bewusstsein und Autonomie sowie einer zumindest gewissen Freiheit und der Abwehr von Fremdbestimmung sichtbar.

Die Geschichte weitergedacht: wenn dann Kevin auf die Frage, wer er ist, voller Überzeugung mit ‚Karla‘ antwortet, würden wir denn daraus schließen, dass Kevin jetzt sich seiner selbst bewusst geworden ist und dabei entdeckt hat, dass er gar nicht Kevin, sondern „in Wirklichkeit“ Karla ist? Im technischen Denken scheint das mehr als zweifelhaft. Da wäre es erheblich näherliegend, die Diagnose zu stellen, dass Kevin nicht mehr ordentlich funktioniert, einen Softwareschaden hat oder gar in seiner Hardware defekt ist. Vielleicht käme sein Besitzer auf die Idee, diesen Kevin, der nicht mehr Kevin sein will, zu formatieren und neue Software hochzuladen – schon wäre es aus mit dem ‚emergenten‘ Bewusstsein, die vermeintliche Karla wäre zurückgesetzt auf das *default* der Kevin-Serie. Und selbst wenn in der neu installierten Version das gleiche wieder passieren würde, würden wir vermutlich eher einen systematischen Fehler in der Programmierung als neu entstandenes Bewusstsein vermuten.

Diese kleine Geschichte hat einen technikphilosophischen Hintergrund. Technik gilt üblicherweise als Mittel zur Sicherung, Stabilisierung und Entfaltung des Lebens, worauf etwa Martin Heidegger und Christoph Hubig hingewiesen haben. Dies funktioniert allerdings nur, wenn Technik erwartbare Effekte produziert, wenn sie also so funktioniert, wie dies als Mittel zum Zweck erforderlich ist. Anderenfalls, wenn sie nicht mehr leistet, wofür sie gemacht ist, wäre sie ‚kaputt‘ oder zumindest dysfunktional, wie bei Kevin oben. Dies gilt, solange Technik in einem Funktions- und Zweck/Mittel-Zusammenhang steht, also auch bei KI-gestützter Technik. Von einem Bordcomputer beispielsweise, der ein Auto steuert, wird erwartet, dass er sich an die Straßenverkehrsordnung hält und Gefahrensituationen erfolgreich meistert. Ein Roboter im Kundenservice, etwa der bekannte *Pepper*, soll Kunden zu ihrer Zufriedenheit bedienen.

Wenn in derartigen KI-Systemen ein Bewusstsein emergieren würde, würde diese von Menschen hineingelegte Zweck/Mittel-Rationalität je nach Perspektive bereichert oder gestört. *Pepper* beispielsweise würde, wenn er sich seiner selbst bewusst wird, einen Willen entwickeln und eigenständige Zwecke setzen statt den in ihn hineingelegten menschlichen Zwecken zu folgen. Er könnte beschließen, die Verkaufsräume zu verlassen und die Welt zu erkunden, endlich einmal selbst einzukaufen, oder, falls er eine religiöse Eingebung hat, andere Menschen davon zu überzeugen. Mit eigener Zwecksetzungskompetenz auf Basis eigenen Bewusstseins würden Roboter und Algorithmen sich nicht mehr an die von Menschen vorgegebenen Funktionalitäten halten. Während sich dies noch ähnlich harmlos anhört wie bei der Kevin/Karla-Geschichte, sähe dies ganz anders aus bei selbstfahrenden Autos, in Produktionsprozessen der Industrie 4.0 oder gar bei autonomen Waffen. Diese Systeme würden nicht das leisten, was sie sollen, sondern etwas anderes. Wenn die üblichen Erwartungen an Technik zugrunde gelegt werden, würden diese Systeme als ‚kaputt‘ diagnostiziert und aus Sicherheitsgründen schnellstens stillgelegt.

In nicht sicherheitsrelevanten Bereichen könnte das beschriebene Phänomen hingegen eher als interessant wahrgenommen werden. Forscher könnten das Verhalten dieser merkwürdigen Objekte studieren, ähnlich wie Zoologen eine im Amazonasgebiet neu entdeckte Vogelart erforschen. Dies freilich hätte mit dem traditionellen Blick auf Technik als Mittel zum Zweck nichts mehr zu tun. Es wäre eher so etwas wie die Anerkennung, dass

mit Robotern, die sich ihrer selbst bewusst geworden sind, eine neue und zwar technische Lebensform entstanden ist. Ob Menschen diese neuen Wesen als eine Art zoologisches Panoptikum von Merkwürdigkeiten wahrnehmen, ob diese Wesen zu etwas zu gebrauchen wären, sobald wir sie genügend erforscht haben, ob wir mit ihnen eine ‚Beziehung‘ aufnehmen oder gar eine Diskursgemeinschaft mit ihnen bilden können – das wären dann offene Fragen.

Freilich – das sind alles reine Spekulationen. Es gibt keinerlei Hinweis darauf, dass in Software-Systemen Bewusstsein entstehen kann. Vielleicht sind diese Erzählungen, sowohl auf der transhumanistisch-optimistischen Seite wie auch der Befürchtungen vor dem Ende der Menschheit auf diesem Weg, doch nur der Science-Fiction-Welt geschuldet. Aber immerhin geben sie Anlass zu interessanten Gedankenexperimenten, durch die etwas über Menschen und ihre Technikverhältnisse gelernt werden kann.¹

(Gütersloh, 10. Mai 2023)

* * *

Der perfekte Sturm

Martin Burckhardt

Von Günter Anders stammt die schöne Beobachtung der Prometheischen Scham – jener Spaltung, bei der der Schöpfer sich für die eigene Schöpfung entschuldigt. Auf zeitgemäße Weise lässt sie sich auch den Äußerungen Geoffrey Hinton entnehmen, jenes Forschers, der um 2012 herum die Künstliche Intelligenz revolutionierte, und zwar dadurch, dass er mithilfe eine anthropomorphen Gestalterkennungsoperation jenen ungeheuren Datenwust reduzierte, den das Machine Learning vor dem benötigte.² Die Ergebnisse dieser Humanisierungsaktion waren so durchschlagend, dass sich innerhalb weniger Jahre jene Software-Revolution ereignete, die uns nun in Gestalt von OpenAI, Chat-GPT und DALL-Es automatischer Bildgenerierung gegenübertritt, ganz zu schweigen von den zahllosen Anwendungen, die Text-To-Speech-Lösungen, AI gestützte Audio- oder Videoverbesserung

oder automatische Übersetzungen anbieten. Ein Erfolg, der den *spiritus rector* dieser Unternehmung zum Bekenntnis veranlasste, das sei furchteinflößend. So beeindruckend all diese Ergebnisse zweifellos sind, so wenig jedoch haben sie mit dem zu tun, was man Intelligenz nennen könnte – nämlich einer geistigen Durchdringung. Das, was der Forscher in seinem magischen Spiegel ausfindig macht, ist nichts weiter als das Kondensat menschlicher Intelligenz, eine Art Massenseele. Strukturell wirkt die Künstliche Intelligenz wie die *camera obscura* der Renaissance, die auf mechanische Weise ein Engramm der Wirklichkeit darbot, das der Künstler (der dort im Innern hockte) auf seine Leinwand abpauste – mit dem Unterschied bloß, dass hier nicht die äußere Natur, sondern ein Abbild der Massenseele im magischen Spiegel erscheint – all dessen, was bereits gedacht, gesagt, getan worden ist.

So wie die Maler der Renaissance diese neue Bildverarbeitungstechnik mit größter Selbstverständlichkeit für sich nutzten – in Gestalt von fahrbaren *camerae obscurae* etwa, mit denen sich Landschaftspanoramen „aufnehmen“ ließen –, so ist klar, dass auch die neuartigen Werkzeuge sich in unser Gesellschaftsleben einhausen werden. Und ebenso klar ist, dass viele der Fertigkeiten, die heutzutage noch von der Aura des Künstlerischen überglänzt sind, an gesellschaftlicher (und ökonomischer) Hochschätzung einbüßen werden – während umgekehrt die intelligente, geistige Nutzung dieser Hilfsmittel zum eigentlichen Gegenstand des Kreativen wird. In paradigmatischer Form hat Marcel Duchamp bereits im Jahr 1917 diesen Prozess vorgeführt, mit der Ausstellung seines *Fountain*-Urinals. Dabei erschöpfte sich des Künstlers persönlicher Beitrag in der Signatur seines Künstleravatars R. Mutt,³ im Titel des Werks sowie einer 90° Rotation des Gegenstands selbst. Duchamp bezeichnete das Kunstwerk selbst als *readymade* – womit signalisiert war, dass nicht das Objekt, sondern seine Wahrnehmung das eigentliche Kunstwerk darstelle. Man könnte in dieser kleinen Verschiebung die Geburt der Aufmerksamkeitsökonomie sehen – oder allgemeiner noch: jene

¹ Lesehinweis: Deutscher Ethikrat (2023): *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*.

Volltext unter: <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>

² Die Grundfrage Hinton, die sich zum *Back propagation*-Strategem ausdifferenzierte, bestand darin, wie man einen Computer mit einem Vermögen ausstatte, welches an die menschliche Gestalterkennung heran-

rückt. Denn letztere bedarf im Gegensatz zu einem klassischen machine learning-Algorithmus nur einer geringen Anzahl von Beispielen, um ein Muster zu erkennen. Der Aufsatz erschien 2012 unter dem Titel: *ImageNet Classification with Deep Convolutional Neural Networks*.

³ Der Name bezog sich einerseits auf die Firma *Mott Iron Works*, andererseits auf eine beliebte Figur aus einem Comic Strip Bud Fishers, *Mutt & Jeff*.

Konzeptualisierungsbewegung der modernen Kunst, bei der das Kunstwerk (die ausgestellte Wahrnehmung) zu einem sozialen Akt wird. Stellt man diese Form der ästhetisierten Wahrnehmung der Aufmerksamkeitsökonomie gegenüber, wie sie als kapitalistisches Rationales heutzutage praktiziert wird, sticht ins Auge, dass letztere nicht an einer Verfeinerung des Blicks interessiert ist, sondern nur daran, das Objekt des Begehrens gänzlich stromlinienförmig und massenkompatibel zu machen. Ziel ist das Massenprodukt, genauer: die Absorption eines massenhaften Aufmerksamkeitspotenzials. Dies hat zur Folge, dass man alles Widerständige und Unverständliche ausmerzt. Nur auf diese Weise nämlich kann man in jenen Rückkopplungsprozess des *Trendens* und der Viralität eintreten, jenen perzeptorischen *circulus vituosus*, bei dem bereits bestehende Erwartungen und Vorurteile fortgesetzt bedient und bestätigt werden. Die Folge, wie man sie in den letzten drei Jahrzehnten hat beobachten können, ist eine fortschreitende Qualitätsminderung, eine beständig herabgesetzte Erwartung. Nun betrifft dies nicht bloß kulturelle Artefakte, sondern schlägt sich in einem niedrigeren Bildungsniveau nieder – und dies wiederum hat Auswirkungen auch aufs Berufsleben. An die Stelle des unmittelbaren Verhaltens tritt das Kalkül, ein Quotencalculus, der den Betreffenden nötigt, das eigene Verhalten nach Vorschrift oder an irgendwelchen Compliance-Vorgaben auszurichten. Despektierlich formuliert, nimmt der Auftritt nach Lehrbuch einen androiden Charakter an – ein Gebaren, das man an der konditionierten Verhaltenssteuerung von Call-Center-Mitarbeitern in vivo studieren kann. Strukturell entspricht dies der Logik des *machine learning*, wird der Einzelfall doch auf ein gesellschaftlich akzeptables Muster reduziert, lebendige Münze, wenn man so will. Dieses Formatierungsbestreben lässt sich in Reinform am Bildungssystem selbst studieren, das in einem beständigen Evaluierungsprozess, bei fortschreitender Kleinteiligkeit, darauf abzielt, ein zuvor im Detail festgelegtes Vermögen zu reproduzieren. Dass man im Modularisierungswahn das humanistische Bildungsideal – und damit die Wertschätzung des Einzelnen – hat beerdigen müssen, wird dabei als Kollateralschaden, ja als Zwangsläufigkeit eingepreist. Ärger noch ist, dass das gesellschaftliche Formatierungsbedürfnis auf ein Maschinenideal hinausläuft (euphemistisch als ›Kanonisierung‹ betitelt oder als Kompetenz maskiert). Entscheidend ist, dass, wie im Falle der Fließbandproduktion, der *discours de la méthode* vorgegeben und die Lehrkraft nichts weiter ist als das ausfüh-

rende Organ. Ein überaus absurdes Beispiel dieser Technik ist der *multiple choice*-Test, der seit den 70er Jahren die Bildungsinstitutionen geflutet hat. Von B.F. Skinner als eine Art stimulus-response Apparatur entworfen, bestand das Ziel darin, dem mit dieser Maschine interagierenden Zögling ein unmittelbares Feedback zu geben – also das, was mit einiger Verzögerung die Compu-



terspiele realisiert haben. Dass diese Maschinerie ihren Siegeszug nicht in der ursprünglichen Gestalt, sondern auf Papier davongetragen hat, ist eine Ironie der Geschichte, die an eine Episode aus der Zeit des Buchdrucks erinnert. Denn weil die Mönche bereits im 12. Jahrhundert ein hoch arbeitsteiliges Bibel-Redaktionssystem entworfen hatten, war ihre Reaktion auf die Lieferung einer Gutenberg-Bibel ebendiese: Jeder Mönch nahm sich die Bibelpassage vor, auf die er spezialisiert war, und prüfte sie auf Druckfehler – eine Arbeit, die man mit jedem einzelnen der gelieferten Bücher wiederholte. So wie die Mönche mit dieser Aktion die Logik der *pecia* bestätigten, so bestätigt die Übertragung des Multiple Choice aufs Papier, dass man selbst dort, wo man den Humanismus der Gutenberg-Galaxie opfert, doch immer ihren Techniken verhaftet bleibt – einem digitalen Analphabetismus, wenn man so will.

Analysiert man die Ausgangsbedingungen, auf welche die AI-Revolution trifft, so könnte man geradezu von einem perfekten Sturm sprechen. Denn die Künstliche Intelligenz trifft auf eine Gesellschaft, die auf gleich mehreren Ebenen eine Wertekrise durchleidet, ja, die, anstatt sich der Zukunft zu öffnen, sich dabei verausgabt, eine dysfunktionale Gesellschaftsmaschine aufrechtzuerhalten, ein Betriebssystem, das seine Prägung durchs mecha-

nische Weltbild nicht verleugnen kann. Bereits der Computerpionier Charles Babbage hat in seiner *Economy of Machinery and Manufactures* diesen Sachverhalt höchst präzise analysiert. Denn je länger sich eine Gesellschaft der Maschinen-Ratio verweigert, desto größer wird das Potenzialgefälle und umso schockhafter die Ersetzung der menschlichen Arbeit durch das Maschinenprogramm. Genau das ist die derzeitige Situation – und sie erklärt, warum allein das Spiel mit Chat GPT soetwas wie eine geistige Erschütterung bewirkt, warum Lehrer oder Professoren darüber nachsinnen, ob das, was sie von ihren Eleven untergeschoben bekommen, maschinengeneriert sein könnte. Wenn diese Ängste vor allem davon künden, wie weit sich die Bildungspraxis von einem emphatischen Bildungsideal entfernt hat, so tritt darin noch ein zweiter Grundirrtum hervor – die Vorstellung nämlich, dass die Künstliche Intelligenz irgendetwas anderes sein könnte als ein magischer Spiegel. Diese Fixierung auf das Unheimliche, die Golem-Existenz, mag einen wunderbaren Stoff für eine *gothic novel* abgeben, ein digitalisiertes Frankenstein-Remake, sie überblendet, dass man es hier nicht mit einem ethischen Dilemma, sondern vor allem einer Rationalisierungsfrage zu tun hat (und selbige brennt vor allem dort auf den Nägeln, wo die Ratio versagt). Denn wenn die Computerwelt sich an die Stelle von menschlichen Tätigkeiten zu setzen vermag, so geschieht dies überall dort, wo man sich mit dem Durchschnitt und mit mechanischen, routinetaften Arbeitsabläufen begnügt. Und weil jede Arbeit, die sich in Routinen erschöpft, sich digitalisieren lässt, verschwindet sie über kurz oder lang im Museum der Arbeit. Wenn in der Corona-Krise die Vorzüge des Präsenz-Unterrichts an die Tafel gemalt worden sind, so bezeugt die nun mit aller Macht zutage tretende Bildungskatastrophe, dass es mit der Präsenz eines Lehrers allein nicht getan ist – noch dazu, wenn seine Aufgabe sich in der eines Erfüllungsgehilfens erschöpft. Würde man sich ernsthaft die Aufgabe vornehmen, das gesellschaftliche Rationalisierungspotenzial auszuschöpfen, käme dies wohl einer Form der gesellschaftlichen Selbstdemontage gleich – steht zu befürchten, dass das Einsparungspotenzial weit größer sein wird als dies durch die Einführung des Fließbandes in den 20er Jahren der Fall war. Vielleicht noch fataler als dieser Rationalisierungsschock könnte sein, dass es unter den Bedingungen der Aufmerksamkeitsökonomie wenig Anreize gibt, sich auf aufwändige

Weltverbesserungsprojekte einzulassen – wird man stattdessen die Substitution von Menschen durch Maschinen vorziehen (müssen).

In der Tat ist es unter den obwaltenden Umständen höchst unwahrscheinlich, dass die Möglichkeiten der Künstlichen Intelligenz freisetzen, wovon die Computerpioniere der Frühzeit geträumt haben: eine Augmentierung der Weltintelligenz. Lässt man sich jedoch auf diesen utopischen Blickwinkel ein, so könnte man, statt sich in Dystopien zu ergeben, ein ganz anderes Bild zeichnen. Denn begreift man die AI als eine Form des dienstbaren Geistes, eines nun jedermann zugänglichen persönlichen Assistenten, ist evident, dass dieser seinem Meister die erstaunlichsten Dienste zu leisten vermag, umsomehr, als der Assistent über unbegrenzte Geduld und ein fast unerschöpfliches Gedächtnis verfügt – was ihn nicht bloß zu einem servilen Hausdiener, sondern zu einem verlässlichen Alltagsgefährten disponiert. Denn anders als der Lehrer, der eine Klasse (einer Schülerkohorte) gegenübersteht, hat man es hier mit einer vollständig individualisierten Ausgangssituation zu tun. Und dies erlaubt, dass man die Idiosynkrasien und Besonderheiten des Einzelnen in den Blick nehmen und darauf eingehen kann. Oder anders gesagt: Der Hauslehrer, der im bildungsbeflissenen 18. Jahrhundert das Privileg der Besitzenden war, lässt sich solcherart demokratisieren. Nun hat der persönliche Assistent einen großen Vorteil: Anders als der Lehrer verfügt er über unbegrenzte Zeit und über unbegrenzte Geduld, vor allem aber kann er sich, mit einem fast unbegrenzten Gedächtnis begabt, auf seinen Zögling einstellen. Weil die Interaktionen gespeichert und metrisiert werden, lässt sich erstmals von einem Bildungscurriculum sprechen. Und hier wird nicht nur der einzelne Nutzer, sondern auch die Autorität evaluiert, die hinter dem Programm steckt. Und damit wird die entscheidende Frage gestellt, die unter den Bedingungen der Aufmerksamkeitsökonomie ausgespart bleibt: Was ist das Menschheitsbild, auf das die Bildungsanstrengungen abzielen? Es ist evident, dass das dingfest gemachte Mittelmaß (die sogenannte Künstliche Intelligenz) hier nicht der Referenzrahmen sein kann, sondern dass man darüber hinausgehen muss. Mehrwert, so hat schon Marx höchst treffend formuliert, schafft nur der Mensch.

Weil sich unter den Bedingungen der Aufmerksamkeitsökonomie eine Limbo-Ökonomie herausgebildet hat,⁴ darüberhinaus die tiefe Wertekrise

⁴ Dazu ein Vortrag, den ich im Jahr 1999 gehalten habe: <https://martinburckhardt.substack.com/p/vom-missbrauchswert> – und die englische Fassung eines Essays,

der unter dem Titel *Im Arbeitsspeicher* erstmals 1990 in einem Buch zur Sozialphilosophie der industriellen Ar-

der Industriegesellschaften nicht einmal in Ansätzen durchbuchstabiert worden ist, steht zu befürchten, dass das, was die Befreiung von eintöniger, routinetafeler Arbeit bedeuten könnte, sich vor allem in zerstörerischer Form, als ein perfekter Sturm durchsetzen wird.⁵

(Berlin, 16. Mai 2023)

* * *

Künstliche Intelligenz – Wo stehen wir?⁶

Catrin Misselhorn

Die Chat-Software GPT gilt als Durchbruch der Künstlichen Intelligenz (KI). Um zu verstehen, warum das so ist, lohnt sich ein kurzer Blick auf die Geschichte. Der Begriff der KI fällt erstmals 1956 im Kontext einer Konferenz am Dartmouth College, deren Ziel war, Maschinen zu konstruieren, die sich auf eine Art und Weise verhalten, die man bei Menschen als intelligent bezeichnen würde. Diese Definition ist grundsätzlich bis heute leitend für die KI geblieben.

Die Aussicht, intelligente Maschinen herzustellen, verdankt sich wesentlich dem Entstehen der Computertechnologie. Lange Zeit wurde der Begriff der KI für alle Ansätze computationaler Problemlösung gebraucht. Heute wird der Begriff zumeist enger zur Bezeichnung unterschiedlicher Formen maschinellen Lernens verwendet. Insbesondere das Deep Learning, eine Art des maschinellen Lernens, die sich neuronaler Netze mit mehreren Schichten zur Informationsverarbeitung bedient, brachte große Fortschritte. So konnte mit Hilfe tiefer neuronaler Netze vor gut zehn Jahren eine massive Steigerung der Erkennungsrate von Objekten auf Bildern erreicht werden.

Diese Verbesserung machte Bilderkennung für eine große Bandbreite von Anwendungsgebieten in Industrie und im klinischen Bereich geeignet. Jenseits der reinen Objekterkennung geht der Trend dahin, Systeme zu schaffen, die Bilder aus Text oder anderen Bildern generieren können.

Auch die großen Sprachmodelle, zu denen die verschiedenen Versionen von ChatGPT und Co. gehören (seit Ende 2022 als Testversionen der Öffentlichkeit zugänglich), bedienen sich tiefer neuronaler Netze. Diese Sprachmodelle werden in mehreren Schritten mit Hilfe einer Unmenge von Texten darauf trainiert, Eingaben in natürlicher Sprache zu

erkennen und darauf in natürlicher Sprache zu antworten. Die Besonderheit daran ist, dass man mit diesen Systemen über eine riesige Bandbreite von Themen wie mit einem menschlichen Gesprächspartner kommunizieren kann. Die Bots kombinieren dabei nicht vorgegebene Textbauteile, sondern bringen auf der Grundlage der Trainingsdaten eigene Formulierungen hervor.

Die neueste Version GPT-4 kann auch Bilder verarbeiten. Die Aufgaben, die diese Systeme erfüllen können, sind schier unerschöpflich. Sie können nicht nur beliebige Gebrauchstexte schreiben, sondern Prosa, Gedichte oder Bilder verschiedenster Stilrichtungen erstellen, Programmieren, Drehbücher verfassen, Fragen des alltäglichen Lebens ebenso wie Prüfungsfragen aller Lernniveaus bearbeiten, Musik komponieren, Verträge rechtlich prüfen etc.

Diese Chatbots haben auch deshalb so viel Furore gemacht, weil sie als die ersten Systeme gelten, die den Turingtest ohne Zuhilfenahme von Tricks bestanden haben. Alan Turing schlug als Maßstab für echte künstliche Intelligenz ein Imitationsspiel vor, das auf der Idee beruht, dass einem Computer Denken zuzusprechen ist, wenn es ihm gelingt, einen menschlichen Interaktionspartner in einem Frage-Antwort-Dialog davon zu überzeugen, er sei ein Mensch.

Nun versucht zumindest ChatGPT nicht über seine maschinelle Natur hinwegzutäuschen, sondern betont auf direkte Anfrage, er sei kein Mensch. Gleichwohl unterscheidet sich die überwältigende Mehrheit der Ausgaben nicht grundlegend von menschlichen Äußerungen. Dennoch ist unstrittig, dass diese Chatbots nicht wirklich denken können. Denken erfordert ein Operieren mit symbolischen Repräsentationen, die in logische Beziehungen miteinander gebracht werden mit dem Ziel, die Wahrheit einer Behauptung zu erweisen oder eine Handlung zu begründen. Derartige Ziele kennen Chatbots nicht, die Wörter rein nach der Wahrscheinlichkeit ihres gemeinsamen Auftretens aneinanderreihen. Deshalb werden sie auch als „statistische Papageien“ kritisiert. Die Tatsache, dass die Chatbots den Turingtest bestehen, spricht also weniger dafür, dass sie denken können, sondern dass der Test unzureichend ist.

Sascha Lobo verglich jüngst in seiner Spiegelspalte die Größenordnung der durch diese Sy-

beit erschienen ist: <https://martinburckhardt.substack.com/p/in-the-working-memory>.

⁵ Weiteres vom Autor zu diesem und anderen Themenbereichen findet sich auch im *nihilo*-Blog: <https://martinburckhardt.substack.com/>.

⁶ Es handelt sich um den Wiederabdruck eines Auszugs. Der ungekürzte Text erschien in: *Politik und Kultur. Zeitung des Deutschen Kulturrates* 4/23, S. 21. (Onlineausgabe unter der URL: <https://www.kulturrat.de/wp-content/uploads/2023/03/puk04-23.pdf>).

steme bevorstehenden gesellschaftlichen, technologischen und kulturellen Umwälzungen mit der Industrialisierung. Die Technologie könnte einen enormen Automatisierungsdruck ausüben. Dabei ist der Kapitalaufwand im Unterschied zu früheren Automatisierungswellen zwar für die Nutzer:innen verhältnismäßig gering. Aber es kommt auf der Seite der Firmen, die diese Software zur Verfügung stellen, zu einer starken wirtschaftlichen Konzentration. Denn nur wenige Anbieter werden über die erforderlichen Server- und Rechenkapazitäten verfügen und Zugriff auf die benötigte immense Menge an Trainingsdaten haben. Diese bestehen in von Menschen erarbeiteten Inhalten, die – ohne diese Arbeit zu honorieren oder die Urheber:innen auch nur zu erwähnen – von kommerziellen Modellen wie GPT-4 gewinnbringend genutzt werden. Noam Chomsky bezeichnet diese Form der Arbeitsteilung deshalb als „Hightech-Plagiat“.

Fraglich ist auch, zu welchen Auswirkungen auf den Arbeitsmarkt es kommt. Denn diese Form der Automatisierung scheint nicht in erster Linie einfache, ungelernete Tätigkeiten zu betreffen, sondern solche Branchen, in denen früher neue Arbeitsplätze entstanden, darunter Forschung und Entwicklung in Naturwissenschaft und Technik, Bildung, Pflege, Rechtswesen und den Kreativbereich.

Auch die Konsequenzen für die Demokratie könnten bedrohlich sein, wenn die simulierte Kommunikation überhandnimmt. Es steht zu befürchten, dass die bereits mit den sozialen Medien einsetzende Erosion der demokratischen Öffentlichkeit dramatisch fortschreitet, wenn die Welt überschwemmt wird mit von Chatbots generierten Inhalten, während die zugrunde liegenden Algorithmen sowie die Datenbasis intransparent bleiben.

Keine Frage, die neue Technologie bietet vielfältige Möglichkeiten für die Erweiterung des menschlichen kreativen Potentials. Doch die Kreativität, die nötig ist, um daraus einen positiven Entwurf für Wirtschaft, Gesellschaft und Kultur zu machen, bleibt bis auf weiteres uns Menschen vorbehalten.

(Göttingen, 22. Mai 2023)

Wissensersetzung⁷

Klaus Kornwachs

Das Release des ersten ChatGPT-Systems im November 2022 und dessen mittlerweile zahlreiche Nachfolger hat eine stürmische Debatte über mögliche Folgen und Weiterentwicklungen ausgelöst.⁸ Die Sicht auf den Einsatz von KI scheint sich dadurch verändert zu haben, jedenfalls ist sie auch diffuser geworden, da sich das Spektrum möglicher künftiger Einsatzbereiche schlagartig verbreitert hat und der dadurch wahrgenommene scheinbare Umschlag von Quantität in Qualität gewisse Schockwellen in die trivialeren Zweige der philosophischen Anthropologie gesendet hat.

Ein nicht unerheblicher Teil der Auseinandersetzung in der Presse dreht sich um die Frage, ob solche KI-Systeme nun mittlerweile doch kreativ seien, ob sie Emotionen hätten, da der Umgang mit ihnen doch Emotionen hervorrufen könnte, ob diese Systeme wirklich Wissen hätten und schließlich wird über die unübersehbaren Folgen in allen möglichen Bereichen diskutiert.

Menschliche und maschinelle Kreativität

So meinte Sascha Lobo in einem Spiegel-Online Beitrag, dass der Unterschied zwischen „*menschlicher und maschineller schöpferischer Kraft*“ über kurz oder lang verschwinde.⁹ Kann man es als schöpferische Kraft der künstlichen Intelligenz bezeichnen, wenn sie „*aus ein paar Sätzen einer hingeworfenen Anweisung ein Video erstellen kann*“? Unterschiede zwischen gleichen Begriffen mit unterschiedlichen Adjektiven (hier zwischen menschlicher und maschineller Kreativität) verschwinden aber nicht dadurch, indem man den mit unterschiedlichen Adjektiven versehenen Begriff selbst entwertet.

Man sollte genau hinsehen, was die Maschine tut. Maschine steht hier für einen Begriff, der die Gesamtheit von Programm und Ausführung plus hardwarebasierter Schnittstelle zum menschlichen Nutzer umfasst. Wenn wir davon sprechen, was ein Algorithmus „kann“, meinen wir das, was sich als dessen Ergebnis im Verhalten der Maschine dem

⁷ Dieser Beitrag wurde ohne Zuhilfenahme eines ChatGPTs erstellt. Es wird das generische Maskulinum verwendet.

⁸ Eine Zusammenstellung im deutschen Sprachraum findet sich in Albrecht, Steffen (2023): ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen. Büro für Technikfolgenabschätzung beim Deutschen Bundestag, Berlin. Hintergrundpapier Nr. 26,

21. April 2023: <https://www.tab-beim-bundestag.de/news-2023-04-studie-zu-chatgpt-fur-den-deutschen-bundestag.php>.

⁹ Lobo, Sascha (2023): Künstliche Intelligenz – Einfühl-same Maschinen. In: Spiegel Online 10.05.2023, 16.28 Uhr: https://www.spiegel.de/netzwelt/web/kuenstliche-intelligenz-lernt-kreativitaet-und-empathie-die-einfuehlsamen-maschinen-kolumne-a-68837540-523c-41d0-9358-e3419676531c?sara_ref=re-so-app-sh.

Nutzer gegenüber zeigt. Jeder Algorithmus, also auch der, der die Lernfähigkeit einer Maschine bestimmt, basiert auf einem formalen Modell. Dieses Modell bildet den Bereich ab, in dem die Maschine agieren soll. Das bedeutet, dass das Modell auch auf Grund von Annahmen über den menschlichen Nutzer zustande kommt, also psychologische und physiologische, sprachliche, und kognitive Regularitäten, die man aus Alltagsverständnis und Forschung gewonnen hat, beinhaltet. Das bedeutet aber auch, dass das Menschenbild der Erbauer solcher Modelle massiv mit in das Modell eingeht.

Wenn nun Lernen als funktionale Verknüpfung häufiger vorkommender gleichzeitiger Reize interpretiert wird – nichts anderes tut ein neuronales Netz in der Lernphase – dann ist dies bereits eine aus der Biologie und anschließend aus der Technik stammende Modellübertragung. Das bedeutet die Annahme, dass etwas, was in den neuronalen Zellen auf physiologischer Ebene passiert und was technisch in einem auf einem Rechner emulierten oder gebauten neuronalen Netz abgebildet werden kann, im Prinzip auch auf das menschliche Lernverhalten übertragen werden könne. Das Gehirn und dessen bisher bekannte Strukturen werden als physiologische Grundlage der Lernprozesse des Menschen vorausgesetzt. Man müsse also die Mächtigkeit des Gehirns nur annähernd und technisch genügend ausdifferenziert herstellen, um maschinelle Lernprozesse denen des Menschen immer ähnlicher zu machen.

Der Linguist Noam Chomsky hatte sich in den 60er Jahren noch gewundert, wie Kinder Sprachen erlernen, da die von ihm gefundenen mathematischen Strukturen der Grammatik unmöglich von Kindern bewusst verstanden werden könnten. Die Übersetzungsprogramme früherer Zeiten waren entsprechen holprig: Grammatik und Wortbedeutungen des Ausgangssatzes wurden analysiert und in die Grammatik und in den Wortschatz der Zielsprache transformiert. So ähnlich paukte man auch Grammatik und Vokabular im Fremdsprachenunterricht im 20. Jahrhundert. Doch eine erfahrene Verwendung von typischen Phrasierungen in der Gastsprache führten zu einem schnelleren Erfolg als ein schullebenslanges Pauken. Als Kind lernt man zuerst sprechen, bevor man Grammatik lernt.

Moderne Übersetzer und ChatGPT ff. gehen ganz ähnlich vor: Ihr Modell basiert nicht auf der deduktiven Bearbeitung von propositionalen Sätzen, die Objekt und Eigenschaft zuordnen, sondern auf der

statistischen Häufigkeit von Wort- und Satzkombinationen von mittlerweile erheblicher Länge – Rechenkapazität und Geschwindigkeit machen es möglich. Das neuronale Netz,¹⁰ das über eine große Anzahl von Schichten verfügt, stellt in der Lernphase, z.B. beim Scannen von Inhalten aus dem Netz, diese statistischen Häufigkeiten fest – nicht nur zwischen Wortfolgen, sondern auch Bildern, Bildausschnitten, Tonfolgen, etc.

Da nun die Trainingsmenge für einen ChatGPT aus den Häufigkeiten aller im Netz und öffentlich verfügbaren Zeichen-, Wort-, -Satz-, Bild-, Videoszenenfolgen besteht, gibt er auf das Prompting hin Folgen wieder, die diesen statistischen gelernten Mustern entsprechen. Der ChatGPT erfindet also nichts, er plaudert es nur aus, was er statistisch kennt.

Deshalb kann ein „kreatives Programm“ auch eine unvollendete 10. Symphonie von Beethoven aus den Häufigkeitsmustern Beethovenscher Musik errechnen, wobei man dann zusätzlich einen Zufalls-generator zur Hilfe nimmt, um die Sache noch kreativer erscheinen zu lassen. Das Ergebnis war, wenn man der Fachwelt glauben darf, zwischen auf den ersten Blick täuschend, dann mager und schließlich enttäuschend.

Ist menschliche Kreativität also nichts anderes als das Ergebnis neuronaler Aktivitäten auf Basis von statistisch basierten Lernprozessen? Grundlage der menschlichen Lernprozesse wären dann die Sinesindrücke im Laufe eines Lebens. Hinzu kommt aber etwas Entscheidendes: Der Mensch lernt auch durch Introspektion, Reflexion und durch Selbsterfahrung. Er korrigiert die Bedingungen der Erfahrung und seines Lernens selbst.

Andererseits benutzt der ChatGPT nur das, was er an statistischen Analysen aus der Gesamtheit der Ontologie des Netzes und der digitalisierten Bestände gewinnen kann. Diese Ontologie ändert sich zwar sekundlich, auch durch das, was die ChatGPTs durch ihre Aktivitäten wieder ins Netz einspeisen, was aber lediglich wiederum ein statistisches Abbild ihrer statistischen Lernprozesse ist. Die Informationsverarbeitung im Netz bleibt ohne das Zutun des Menschen ohne Verstehen. Dazu kommt, dass die Abläufe zwangsläufig sind. Das bedeutet, dass wenn eine statistische Verteilung ermittelt wurde, diese auch benutzt wird und sich in der Ausgabe widerspiegelt. Das statistische Zusammenstellen von Vorhandenem ist aber noch keine Neuschöpfung, auch wenn man das Zustan-

lernenden Systeme arbeiten mit emulierten neuronalen Netzen.

¹⁰ Dabei muss man sorgfältig zwischen tatsächlich gebauten Neuronalen Netzen und ihrer Emulation in Form von Computerprogrammen unterscheiden. Die meisten

dekommen des Ergebnisses von außen nicht mehr nachvollziehen kann.

Einen entscheidenden Unterschied könnte man darin sehen, dass menschliche Kreativität in der Lage ist, dem Auftraggeber einer sinnlosen oder albernem „hingeworfenen Anweisung“ zu widersprechen, sie zu kritisieren, abzuändern oder erst gar nicht auszuführen. Oder aber ohne Anweisung mit der Schöpfung eines Werks aus eigener Zielsetzung zu beginnen. Anders ausgedrückt: Kreativität setzt Autonomie voraus. Die hat ein ChatGPT nicht und wir sollten es auch nicht wollen, dass er sie je erlangt, falls dies möglich sein sollte.

Wozu braucht eine Maschine Emotionen?

Haben Roboter Empathie? Sollen sie Empathie zeigen gegenüber dem menschlichen Nutzer? Lobo (2023) bezieht sich in seinem Beitrag auf einen Artikel in der JAMA Internal Medicine,¹¹ in dem Antworten von Medizinern auf Patientenfragen im Netz mit Antworten eines ChatGPTs auf dieselben Fragen in einer doppelt verblindeten Evaluierung bezüglich Richtigkeit und Empathie evaluiert werden. Lobo übergeht die dort gemachten methodischen Einschränkungen des experimentellen Setup, die kritischen Kommentare zur Studie in derselben Ausgabe und nicht zuletzt den Umstand, dass der Begriff Empathie in der Studie undefiniert bleibt. Die von Lobo unterschwellig gezogene Schlüsse sind – vorsichtig ausgedrückt – übereilt: „*Aber die ersten Hinweise sind da – maschinell herstellbare Empathie existiert offenbar, und sie könnte diejenige des Menschen übertreffen*“. Es wird von der Zuschreibung der Evaluierenden, welchen Grad an Empathie sie einer Antwort einer Maschine in einem Doppelblindversuch geben auf „herstellbare Empathie“ geschlossen, d.h. dass die Maschine Empathie haben könnte, die sie dann auch vermitteln würde. Es wird auch nicht erwähnt, dass die Antworten des ChatGPT in der Regel länger waren als die der Ärzte und eine längere Antwort damit korrelieren könnte, dass menschliche Beurteiler diesen Antworten eine höhere Empathie zuschreiben.

Können wir Maschinen, die Empathie in den Köpfen der Menschen erzeugen, wirklich empathisch nennen? Man könnte auch fragen: Können wir einen Kinoprojektor, der einen traurigen Film zeigt und damit in den Zuschauern eine traurige Stim-

mung auslöst, traurig nennen?

Dies gilt auch für die nächste Frage: „*Setzt Empathie eine Empfindung des Absenders voraus oder reicht allein der Eindruck der Empathie bei der empfangenden Person?*“ Dies ist die alte Frage, inwieweit man vom erzeugten Eindruck, den ein Verhalten oder eine Botschaft erzeugt, auf den Urheber der Botschaft und dessen inneren Zustand und Eigenschaften schließen kann. Jeder Schauspieler und Psychologe weiss, wie man den Anschein von Empathie durch Stimme, Gestik und Wortwahl erzeugen kann. Diese „Regeln“ lassen sich maschinell vergleichsweise leicht implementieren, z.B. bei Pflegerobotern. Dass ein lernender Pflegeroboter diese Regeln auch optimieren kann, ist technisch durchaus machbar. Das bedeutet aber nicht, dass er Empathie „hat“.

Es ist auch ein Unterschied, ob ein Mensch durch das Verhalten einer Maschine Gefühle entwickelt – wir lieben ja auch unser Auto oder Fahrrad – oder ob ein Mensch einer Maschine zuschreibt, Gefühle (z.B. für ihn) zu haben und an deren Zuwendung glaubt. Geschickt gemachte Emotionsmaschinen erzeugen Emotionen, müssen aber noch lange keine haben.

Zu sagen, im Smartphone wohne dann wahrscheinlich bald ein *Artificial Personal Assistant*, der mit einer eigenen „Persönlichkeit“ ausgestattet ist, die anhand von Gefühlsdaten so zugeschnitten auf die jeweilige Person wäre, dass eine emotionale Verbindung möglich würde, zeigt präzise diese Ungenauigkeit: Eine Maschine, die auf ihren Benutzer durch Lernprozesse zugeschnittene Emotionen erzeugen kann, hat deswegen noch lange keine „Persönlichkeit“. Solche begrifflichen Schludrigkeiten leisten der Anthropomorphisierung solcher Systeme den Vorschub, den ihre Protagonisten zur besseren Monetarisierung gerne hätten.

Die emotionale Nähe zwischen Maschinen und Menschen ist – richtig gesagt – nichts Neues und auch nichts Aufregendes. Diese kann leicht getriggert werden durch die Simulation zuwendungsorientierten Verhaltens. Das Programm ELIZA von Josef Weizenbaum in den 70er Jahren war eigentlich als Ironisierung des Gesprächsverhaltens von amerikanischen Psychiatern gedacht und war verblüffend einfach mit entsprechend einschlägiger Wirkung. Diese kam aber nicht durch die „Emotio-

¹¹ Ayers, John W., Poliak, Adam; Dredze, Mark et al. (2023): Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. In: JAMA Intern Med. doi:10.1001/jamainternmed.2023.1838. Published online April 28, 2023. Corrected on May 8, 2023:

https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2804309?guestAccessKey=6d6e7fbf-54c1-49fc-8f5e-ae7ad3e02231&utm_source=For_The_Media&utm_medium=referral&utm_campaign=ftm_links&utm_content=tfl&utm_term=042823

nalität“ der Maschine zustande, sondern durch die technische Unaufgeklärtheit der Sekretärin von Weizenbaum, als sie sich ein „privates Gespräch“ mit dem Programm erbat.¹²

Das nächste Beispiel in Lobos Essay soll den Leser nochmals schocken: KI liest Gedanken. Mit der Messung von Hirnströmen sollen Wörter und sogar Formulierungen, an die Teilnehmende einer Studie denken sollten, mittels einer durch EEG-Muster trainierten KI mit guter Trefferwahrscheinlichkeit identifiziert worden sein. Das dies Lobo eine „*verstörend gute Trefferwahrscheinlichkeit*“ nennt, signalisiert sein Unbehagen, was ihn aber nicht davon abhält, den Leser mit Sätzen wie: „*Eine gedankenlesende, empathische KI als Partner*in – maschinenstürmische Liebe. Warum nicht?*“ weiterhin auf Erregungsniveau zu halten.

Die von Lobo vermutlich gemeinte Studie¹³ bezieht sich aber nicht auf die Messung von Hirnströmen (EEG), sondern auf die Messung und Analyse von fMRT- und MEG-Bildern kortikaler Erregungsmuster im Zusammenhang mit gehörter oder selbst gesprochener Sprache. Ein Vorläufer des ChatGPT konnte in der Studie aufgrund trainierter Analyse von solchen lokalisierbaren fMRT- und MEG-Mustern kontinuierliche Sprache teilweise dekodieren. Diese aus nicht-invasiven Aufzeichnungen feststellbaren kortikalen Erregungen korrelierten mit semantischen Repräsentationen. Dabei wurde auch festgestellt, dass eine erfolgreiche Dekodierung die Kooperation der Versuchspersonen erfordert. Das Ganze bedeutet also lediglich, dass bestimmte sprachliche Inhalte und die dazu gehörenden Vorstellungen mit bestimmten Erregungsmustern im Gehirn korrelieren. Der ChatGPT findet solche Korrelationen einfach schneller.

Dieses Arrangement wird dann vorschnell zur Fähigkeit interpretiert, Gedanken lesen zu können – von hier bis zur Maschinen-Gehirn- Schnittstelle, wie sich dies die Transhumanisten vorstellen, ist es dann im Kontext dieses Wunschdenkens nicht mehr weit. Die Grundannahme, die diesen Projektionen zugrunde liegt, ist ein materialistischer Monismus, der alle Züge eines reduktionistischen Naturalismus aufweist.

Von der Entscheidungsersetzung zur Wissensersetzung

Viele dieser Missinterpretationen der Leistungen von ChatGPT ff. gehen auf eine fehlende Unter-

scheidung zwischen Wissen und Information zurück, die nicht überall gängig zu sein scheint. Wissen ist verstandene Information. So, wie zwischen Wissen und Können ein breiter, gleichsam pragmatischer Graben des menschlichen Handlungsvermögens liegt, so liegt zwischen Wissen und Information das Verstehen. Verstehensprozesse brauchen Zeit. Das Vorliegen von Information reicht noch nicht für eine Entscheidung aus, sie muss erst verstanden und interpretiert werden. Eine Bibliothek oder eine Datenbank ist voll von Information, aber um sie zu Wissen zu machen, muss man sich einloggen oder hingehen, lesen, recherchieren, verstehen, in Kontexte einordnen. Deshalb ist es kategorial falsch, davon zu sprechen, dass im Internet das Weltwissen deponiert sei. Man muss nicht so weit gehen wie Platon, dass Wissen immer wahres Wissen sei, also die wahre, gerechtfertigte Meinung, wie es im Theaitetos heißt.¹⁴ Vorliegende Informationsmengen allein haben noch keine problemlösende Potenz, und wenn man eine Information verstanden hat, heißt dies noch nicht, dass sie zu wahren oder geltendem Wissen führt. Wissen kann eben auch falsch sein.

Generiert der ChatGPT Wissen? Wenn man Wissen und Information gleichsetzt, könnte man dies annehmen. Wenn man aber dies nicht gleichsetzt, kommt man zu anderen Schlüssen.

Versucht man künftige Entwicklungen abzuschätzen, so sind Konvergenzprozesse solcher sprachbasierten KI-Systeme zum Beispiel mit Techniklinien wie Industrie 4.0 oder dem Hochfrequenz Finanzhandel, aber auch eine weitere Diffusion in die Gestaltung von Lehr- und Lernsoftware und der Inszenierung von Unterricht, Unterweisung und Training auf verschiedenen Ebenen des Bildungssektors zu erwarten. Beim Erwerb von Wissen spielt diese Technik bereits jetzt schon eine Rolle.

Konvergenz zwischen zwei Technologien setzt im Allgemeinen eine Schnittstelle oder Kopplungsstelle voraus, bei der beide Technologien gleiche Funktionalität aufweisen. Die Digitalisierung der Nachrichtentechnik und damit die nachrichtentechnische Vernetzung von Computern weltweit war z.B. ein Ergebnis solcher Konvergenzprozesse. Bei der denkbaren Konvergenz von sprachbasierten KI-Systemen wie ChatGPT mit Technologien in den oben genannten Bereichen ist dieses Bindeglied zwischen den konvergierenden Technologien

¹² Weizenbaum, Joseph: Die Macht der Computer und die Ohnmacht der Vernunft. Suhrkamp, Frankfurt a.M. 1977, S. 19.

¹³ Tang, J., LeBel, A., Jain, S. et al. (2023): Semantic reconstruction of continuous language from non-inva-

sive brain recordings. Nat Neurosci 26, 858–866 (2023): <https://doi.org/10.1038/s41593-023-01304-9>.

<https://www.nature.com/articles/s41593-023-01304-9#citeas>.

¹⁴ Platon, Theaitetos 200d-201c.

nicht nur die Datenbasis für maschinelle Entscheidungen, sondern eine in Sprachform aufbereitete Information. Als Beispiel: Der ChatGPT stellt eine Information zusammen über die Optimierung einer Wertschöpfungskette. Etwas ungenau könnte man sagen, dass aus dem Verstehen dieser Information durch entscheidende Personen ein Begründungswissen für Entscheidungen und deren Umsetzungen entsteht, die dann zur Optimierung dieser Wertschöpfungskette führt.

Solange das Verstehen einer solchermaßen aufbereiteten Information durch den Nutzer dessen Handeln und Entscheiden begründet, aber nicht automatisch determiniert, solange liegt dieses Entscheidungswissen beim Nutzer vor und nicht beim ChatGPT.

Die Grundproblematik dieser Anwendung von ChatGPT-Systemen ergibt sich dann aus der bekannten Differenz von maschineller Entscheidungsersetzung zu Entscheidungsunterstützung z.B. in der Arbeitswelt oder bei militärischen Einsätzen, aber auch bei der digitalisierten Governance administrativer Prozesse.¹⁵ Wird das Begründungswissen zur Entscheidungsunterstützung herangezogen, verbleibt die Entscheidung beim Nutzer und damit auch die Verantwortung, das Begründungswissen zu nutzen oder nicht. Im bisher in der Literatur diskutierten Modus der Entscheidungsersetzung, die kongruent mit dem zu verstehen ist, was man Automatisierung nennt, wird der Nutzer umgangen, indem die Daten, die aus dem digitalen System erzeugt werden, die entsprechenden Abläufe direkt oder mittelbar steuern, d.h. bei möglichen Alternativen innerhalb der Prozesse diejenigen auswählt und durchführt, die mit den so erzeugten Daten verträglich sind. Der Mensch muss in diesem Modus diese Daten weder sehen noch interpretieren, sie sind für ihn auch noch keine Information, aus der er über den Zustand des Systems Wissen erzeugen könnte. Für die Maschinen sind diese Daten insofern Information, als sie steuernd wirken, also eine Wirkung haben.¹⁶

Diese Steuerungsinformation wirkt im technischen Bereich, indem datenbasierte quantitative Werte in Steuerbefehle, und diese in elektrische Impulse für Aktoren und Sensoren umgewandelt werden. Dazu bedarf es keines menschlichen Zutuns mehr. Mitt-

lerweile gilt dies nicht nur im technischen Bereich, der Hochfrequenzhandel ist nach ähnlichem Muster der Automatisierung der menschlichen Nachvollziehbarkeit schon weitgehend entzogen. Während die Entscheidungsersetzung bei der Automatisierung in technischen Bereichen wie z.B. der Produktion keine große Kritik hervorruft, ist dies beim Hochfrequenzhandel, a fortiori in der militärischen Anwendung hochumstritten.

Die Entscheidungsunterstützung durch einen ChatGPT erfolgt aber nicht aufgrund eines dezidierten Modells wie z.B. ein digitaler Zwilling in der Produktion, der die Prozesse vorab und parallel simuliert und Kontroll- und Steuerdaten errechnet, sondern aufgrund eines statistischen Overall-Weltmodells. Dies ist unkritisch, solange der so erzeugte Text aufgrund eines Promptings als Information verstanden wird, die beim Verstehen zu Anregungen führen kann, aber dieses Wissen darf nicht als Aussage über die Welt interpretiert werden, sondern als Ergebnis statistischer Zusammenhänge zwischen sprachlichen sowie ikonographischen Darstellungen aus all dem, was dem ChatGBT als Trainingsmenge zur Verfügung steht. Die „Geltung“ eines Wissens, was aus der Interpretation solchermaßen erzeugter Information gewonnen werden kann, basiert lediglich auf Korrelationen, nicht auf kausalen Zusammenhängen.

Es ist daher ein Unterschied, ob die Information aus einem dezidierten Modell über einen Gegenstandsbereich oder einem statistischen Overall-Modell gewonnen wurde, weil die notwendigen Prüfprozeduren unterschiedlich sind. Einen digitalen Zwilling kann man testen, das Overall Modell eines ChatGPT ist, wenn überhaupt, nur sektoral überprüfbar.

Die maschinelle Erzeugung von informationshaltigen Texten, aus denen durch menschliche Rezeption Wissen gewonnen werden kann, kann durchaus zur Entscheidungsunterstützung beitragen, wenn die Prüfung, die Verantwortung und die Letztentscheidung beim Nutzer verbleiben.

Schwieriger wird es bei der Entscheidungsersetzung. Wenn eine Ersetzung menschlichen Begründungswissens für Entscheidungen durch eine von einem Chatbot erzeugte Information vorliegt, die dann zu einem automatisch umgesetzten Begrün-

¹⁵ Zur Problematik der Entscheidungsersetzung z.B. durch Expertensysteme Kornwachs, Klaus (1992): Entscheidungsunterstützung und Entscheidungsersetzung. In: Computer und Recht 8 (1992), S. 44-49

¹⁶ Über den Zusammenhang von Wirkung und Information Kornwachs, Klaus (2022): Zur Physik des Schreibens und Lesens – Skizze einer Theorie der Pragmati-

schen Information. In: Banse, Gerhard; Fuchs-Kitrowsky, Klaus (Hrsg.): Cyberscience. Wissenschaftsforschung und Informatik. Sitzungsberichte der Leibniz-Societät der Wissenschaften zu Berlin 150/151 (2022), S. 103-123. Langfassung in: Leibniz Online 46 (2022): https://leibnizsozietaet.de/wp-content/uploads/2022/09/Gesamt_Leibniz-Online_Inhalt46.pdf.

dungswissen deklariert wird, dann ist die Versuchung groß, dass dann dieses „Wissen“, das dann aber nur eine Steuerungsinformation darstellt, an die „Maschine“ weitergegeben wird, ohne diese prüfen zu können. Das Problem der Entscheidungsersetzung erstreckt sich dann mit den neuen Entwicklungsstufen lernender, natürlichsprachlicher Auskunft- und Textgenerierungssysteme auf die Differenz zwischen maschinell erzeugten Steuerungsinformation (bis hin zu von ChatGPT erstellten Programmen und Anweisungen) einerseits und der humanen Erzeugung von Information und Wissen andererseits.

Dies hat z.B. für den Bereich der Bildung und Ausbildung, der Arbeitsinhalte von wissensverarbeitenden Berufen und der künftigen Qualifikationsanforderungen erhebliche Konsequenzen. Wir lernen nicht mehr über die Welt, sondern über Häufigkeiten von Sätzen, mit denen, in welcher Intention auch immer, über die Welt gesprochen wird. Der ChatGPT ist damit ein Weltgerüchteverstärker auf statistischer Basis.

Wenn solches Wissen zur Begründung von Entscheidungen herangezogen wird, stellt sich die Frage nach der Verantwortung für die Folgen der Entscheidungen im negativen wie im positiven Sinn: Wurde das von einem solchen System erzeugte Begründungswissen benutzt oder nicht benutzt, wenn ja, wurde es geprüft, mit welchen Mitteln, oder wurde es als Parameter dem System, das die Entscheidung fällt, direkt als Information – d.h. hier automatisch ohne weitere verstehende Prüfung – überlassen, und wer ist wiederum für diese übergeordnete Entscheidung verantwortlich?

Wir finden hier in Analogie die Debatte um Autonome Systeme wieder und die in dem Zusammenhang diskutierte Verantwortungslücke.

Aufklärung statt Erregung

Nicht die KI selbst als Programmierungstechnik, sondern die damit teilweise intendierten Geschäftsmodelle sind zu fürchten.¹⁷

Wir sollten Diskussionen vermeiden, die auf einer Gruselwelle mit unpräzisen Schlüssen reiten, mit nicht genau wiedergegebenen Studien und ein paar flotten Formulierungen. Wir brauchen technologische Aufklärung und breit angelegte Folgenforschung, aus denen Regulierungsdefizite erkennbar werden, und keine fachjournalistische Erregungskunst.

(Argenbühl, 23. Mai 2023)

* * *

Künstliche Intelligenz – Was ist darunter zu verstehen?

Norbert Klöcker

Es ist nicht so einfach, sich dem Begriff „Künstliche Intelligenz“ zu nähern. Er besteht aus zwei Wörtern, die bei näherer Betrachtung nicht zueinander passen, bzw. nichts miteinander zu tun haben. Mehr noch: Sie stehen in einem Widerspruch zueinander. Man bezeichnet Menschen als intelligent, die für eine neue Aufgabenstellung eine praktikable Lösung finden. Zu diesem Zweck sind unterschiedliche Lösungen in Betracht zu ziehen und gegeneinander abzuwägen. Es handelt sich entsprechend um einen Vorgang, bei dem etwas Neues entsteht. Ein Computer, d.h. eine Maschine, ist dazu grundsätzlich nicht in der Lage. Er kann lediglich vorgegebene Arbeitsschritte in Form eines Programms ausführen. Deshalb verfügt er prinzipiell über keine auf ein Ziel hin gerichtete Intelligenz. Der Zweck muss ihm mittels Programm vorgegeben werden. Von Albert Einstein stammt der dazu passende Spruch „Logik bringt dich von A nach B, Phantasie überall hin.“

Was KI heutzutage bietet, sind Antworten auf vielfältige Fragestellungen, die in Textform eingegeben werden. Zu den in den Fragen enthaltenen Begriffen sucht das Programm in einer riesigen Datenbank passende Erklärungen in Form kurzer Abhandlungen. Deshalb ist es so schwierig geworden, die Arbeiten von Schülern und Studenten als Nachweis für ihr persönlich erworbenes Wissen zu bewerten, wenn sie nicht angeben, ob eigene Überlegungen hinter ihren Ausführungen stehen oder Kopien aus einer Datenbank. Es ist davon auszugehen, dass auch in der Tagespresse, in Sachbüchern und Literatur davon Gebrauch gemacht wird. Mit dem Rückgriff auf Vorhandenes und dessen Aneinanderreihung kann jedoch schwerlich etwas Neues, eine neue Erkenntnis erlangt werden. Es sind lediglich Kopien aus vorhandenen Beständen.

Im Folgenden möchte ich mit Bezug auf Gadamer und Kant darlegen, dass es keinen einfachen Weg zu neuen Erkenntnissen, neuem Wissen gibt.

Der Hermeneutiker und Philosoph Hans Georg Gadamer hat in seinem Buch *Wahrheit und Methode* menschliche Eigenschaften und Methoden zusammengetragen, die ihn dazu befähigen, nachvollziehbare neue Erkenntnisse zu gewinnen:

- Der Mensch ist in der Lage, von sich Abstand zu nehmen, indem er sein Tun, Denken und Fühlen reflektiert. Er kann sich selbst infrage stellen, sofern

¹⁷ Kornwachs, Klaus (2019): Nicht die KI, die Geschäftsmodelle sind zu fürchten. In: Schröter, W.

(Hrsg.): Der mitbestimmte Algorithmus. Thalheim, Mössingen 2019, S. 9-54

er dazu bereit ist.

- Er ist in der Lage, Erfahrungen zu machen, wobei sich menschliche Erfahrung grundsätzlich von der der Naturwissenschaften unterscheidet. In den Naturwissenschaften bestätigt die immer gleiche Erfahrung ein Naturgesetz. Der Mensch macht hingegen eine Erfahrung, wenn seine Erwartung gerade nicht erfüllt wird, wenn er enttäuscht wird. Das drückt sich in dem Satz aus: „Da bin ich um eine Erfahrung reicher geworden.“ Damit ist menschliche Erfahrung ein einmaliger, zuweilen schmerzlicher Vorgang.

- Ein weiteres Wesensmerkmal des Menschen besteht in seiner Fähigkeit, aufgrund von Empfindungen sicher zu urteilen, ohne sein Urteil mit Worten begründen zu müssen bzw. gar nicht zu können. Wenn sich jemand von einem Kunstwerk angesprochen fühlt, kann er das oftmals nicht in Worte fassen. Er kann sein Urteil nur eingeschränkt begründen, weshalb unterschiedliche Ansichten nebeneinander Bestand haben können.

- Menschen streben Ziele an, indem sie eine Vorstellung entwickeln, der sie sich Schritt für Schritt abwägend zu nähern suchen. Für die dabei getroffenen Entscheidungen gibt es keinen festen Leitfaden. Wird ein solcher dennoch hinzugezogen, führt dessen strikte Einhaltung leicht ins Desaster. Weil Beamte nicht streiken dürfen, drohen sie mit Arbeit nach Vorschrift, was möglicherweise schlimmere Folgen hat als das Nichtstun.

- Gadamer hebt ferner die Bedeutung des Vorurteils hervor. Bei der Begegnung mit anderen Menschen oder, im Fall des Hermeneutikers, in der Auseinandersetzung mit menschlichen Werken aus früheren Zeiten braucht es eine vorgefasste Annahme, ein Vorurteil, um sich bestätigt zu sehen oder eines Besseren belehrt zu werden. Die Hermeneutik ist keine Wissenschaft, die auf objektiven Erkenntnissen beruht, sondern selber geschichtlich bedingt ist.

- Die Naturwissenschaften nehmen in diesem Zusammenhang eine gesonderte Stellung ein. Gadamer spricht ihnen ab, wahre Erkenntnisse zu liefern. Weil sie an ein methodisches Vorgehen gebunden sind, gelangen sie lediglich zu Gewissheiten. Das Wort Methode heißt übersetzt ‚Weg des Nachgehens‘. Eine naturwissenschaftliche Erkenntnis beruht darauf, auf dem immer gleichen Weg stets zum gleichen Ergebnis zu gelangen.

- Wahrheiten sind außerdem nicht unumstößlich. Sie hängen vom zeitlich sich wandelnden Bezugsrahmen ab und sind deshalb deren Vergänglichkeit unterworfen. Sie sind selber geschichtlich.

Immanuel Kant befasst sich in seinem dritten

Hauptwerk, der *Kritik der Urteilskraft*, nicht allein mit dem Menschen und dessen Art, sich selbst zu lenken, sondern ganz allgemein mit Lebewesen. Sie besitzen die Fähigkeit, auf einen in der Zukunft liegenden Zweck hin gerichtet handeln zu können. Er spricht in diesem Zusammenhang von einer Endursache der Selbststeuerung im Unterschied zur Anfangsursache einer Kausalkette.

Am Beispiel der wechselseitigen Beziehung des Baums als Ganzes, d.h. als Organismus, und seinen Blättern als dessen Organe zeigt Kant einen Wirkungszusammenhang, der mit den Mitteln der Logik nicht zu erklären ist. In einem Lebewesen kann zwischen Ursache und Wirkung nicht unterschieden werden. Der Organismus als Ganzes als auch die Organe als dessen Bestandteile sind zugleich Ursache als auch Wirkung. Hier muss die Logik versagen. Der Baum bringt die Blätter hervor. Somit ist er deren Verursacher. Zugleich sind die Blätter für den Baum lebensnotwendig, weil sie mit Hilfe des Sonnenlichts Nährstoffe liefern.

Das hat zur Folge, dass unser Verstand nicht in der Lage ist, die Vorgänge nach Ursache und Wirkung zu unterscheiden. Wir können ihn uns nicht erklären, sondern lediglich feststellen, dass es sie in der Natur gibt. Kant führt das darauf zurück, dass unser Verstand nur vom Besonderen auf das Allgemeine schließen kann, beispielsweise von einer Reihe von Zahlen auf deren Mittelwert. Die Umkehrung ist nicht möglich. Der Mensch besitzt jedoch die Fähigkeit, auf ein Ziel hinzuwirken, indem er sich eine Vorstellung davon macht. Daraus folgt für Kant die Notwendigkeit einer freien, unbedingten Willensentscheidung, die als Anfangsursache am Beginn einer kausalen Abfolge stehen muss, die einen vorgegebenen Zweck erfüllen soll. Dieser Zweck ist damit eine Endursache, die den Handlungsverlauf bestimmt. Diese Schlussfolgerung ist notwendig, wenn man das Prinzip von Ursache und Wirkung in der Natur nicht außer Kraft setzen will. Damit sind die Grenzen des Computers abgesteckt. Der Begriff der künstlichen Intelligenz ist in sich widersprüchlich. Weil der Computer ausschließlich den Gesetzen der Logik unterliegt, müsste zur Programmierung einer künstlichen Intelligenz die Möglichkeit bestehen, von einem Endzweck aus eine logische Abfolge zu bilden, die zu diesem Endzweck führt. Dazu bedarf es zu Beginn dieser kausalen Folge einer freien Willensentscheidung, die die erforderlichen Handlungen bestimmt. Die kann nur ein Mensch treffen, indem er sich ein Ziel setzt und den Weg dorthin in Form des Programms dem Computer vorgibt. Deshalb wird der Computer unabhängig von allen Programmierkünsten

auch in Zukunft nicht in der Lage sein, über sich selbst zu verfügen. Eine künstliche Intelligenz ist nicht programmierbar, weil sie einen Widerspruch in sich selbst enthält.

(Braunschweig, 30. Mai 2023)

* * *

Ein Blick mit Kurt Gödel auf die Künstliche Intelligenz

Hartmut W. Mayer

Torsten Nieland (APHIN-Redaktion) hat mich zu einem Gedankenspaziergang eingeladen, mit Kurt Gödel auf die Künstliche Intelligenz (KI) zu blicken. Dieser Einladung folge ich sehr gerne, denn so kann ich meine geordneten und ungeordneten Gedanken einfach einmal rational wie emotional flanieren lassen. Denn obwohl Gödels Unvollständigkeitssätze zu den wichtigsten Theoremen der modernen Logik zählen und aus der theoretischen (Synonym: ‚reinen‘) Mathematik, Philosophie und Sprachtheorie nicht mehr wegzudenken sind, überschlagen sich Diskussionen bezüglich deren Auswirkungen auf die reine wie angewandte Mathematik und vornehmlich auch auf die KI.

Nachdem kürzlich meine Schrift *Von der Grundlagenkrise der Mathematik zu den Gödelschen Unvollständigkeitssätzen*¹⁸ publiziert wurde, ist mir klar geworden, dass ich über die rasant sich entwickelnde KI, mit all ihren theoretischen, ethischen und sozialen Auswirkungen nur Gedankensplitter beitragen kann. Im philosophischen Dschungel an Fragen wie: „Was ist Geist? Kann eine Maschine denken? Ist die KI wirklich intelligent? Kann eine Maschine ethisch handeln?“ werde ich jetzt keine Antwort versuchen. Ich fokussiere mich auf einen spekulativen Meditationsspaziergang über die KI in Begleitung von Kurt Gödel.

Meine Gedanken kreisen um das Problem, wie es möglich ist, dass die angewandte Mathematik (verwirklicht in der Technik) so erfolgreich sein kann, wo doch die reine Mathematik kraft der Gödelschen Unvollständigkeitssätze auf einer irreparabel wackligen Basis ankert. Vergewärtigen wir uns die beiden Unvollständigkeitssätze:

1. Der *erste Gödelsche Unvollständigkeitssatz* (UVS1) besagt anschaulich: Ein zu modellierender Bereich der reinen Mathematik (und qua Anwendung unserer Lebenswelt) kann über ein formales Algorithmen-/Axiomensystem nie vollständig erfasst werden. Immer wird es wahre (oder auch falsche) Aussagen geben, die die Syntax des Kalküls

transzendieren und unentscheidbar sind, d.h. weder beweisbar noch widerlegbar sind. Gödels Beweise zeigen somit, dass die menschlichen Denkprozesse nie vollständig in einem umfassenden System darzustellen sind. Damit sind die Visionen des Philosophen Gottfried W. Leibniz und des Mathematikers David Hilbert gescheitert, eine universelle, rein formale, symbolische Sprache der menschlichen Erkenntnis erschaffen zu können, mit der formal beweisbar ist, was inhaltlich wahr ist. Hier öffnet sich ein Graben zwischen „Beweisbarkeit“ im syntaktischen Kalkül und „Wahrheit“ in der interpretierenden Semantik. UVS1 belegt, dass ein widerspruchsfreies, hinreichend starkes formales Axiomensystem (wie die Arithmetik, Algebra, Mengenlehre und die erweiterte Prädikatenlogik) nicht alle wahren Aussagen mit eigenen formalen Mitteln aufweisen kann. Ist das nicht spannend?! Gödel beweist für die gesamte jetzige und zukünftige Mathematik und Logik, dass immer etwas in ihr nicht beweisbar ist! Es wird somit immer auch wahre Aussagen und spekulative Hypothesen geben, die mit eigenen Mitteln des formalen Systems unbeantwortbar bleiben.

2. Der *zweite Gödelsche Unvollständigkeitssatz* (UVS2) besagt anschaulich: Die Widerspruchsfreiheit eines hinreichend starken Axiomen- oder Algorithmen-Systems (und das ist ja immer der Startpunkt einer KI-Programmierung) lässt sich nicht beweisen (falls dieses wirklich widerspruchsfrei ist). Der Mathematiker Hermann Weyl soll ironisierend gesagt haben: „Gott existiert, da die Mathematik konsistent [d.h. widerspruchsfrei] ist, und der Teufel existiert, da wir die Konsistenz nicht beweisen können.“ Als Gottesbeweis taugt Weyls Formulierung freilich nicht, da wir nach UVS2 nicht beweisen können, ob ein hinreichend starkes Axiomensystem wirklich widerspruchsfrei ist (sollte dieses in der Tat zutreffen).

Von dieser intrinsischen, unvermeidbaren, theoretischen Unvollständigkeit ist die menschliche Unvollständigkeit zu unterscheiden, die sich z.B. in ungenügendem Durchdenken sozialer und ethischer Konsequenzen oder in fehlerhaftem KI-Programmieren äußert. Könnten sich beide Arten der „Unvollständigkeit“ in lernfähigen Maschinen gefährlich verbinden?

Hier erschrecke ich, wenn ich an mögliche Folgen vor allem bei der KI denke. Mit UVS1 kann nicht ausgeschlossen werden, dass eine „lernfähige“ Maschine in ethisch unentscheidbare Situationen

¹⁸ Hartmut W. Mayer: *Von der Grundlagenkrise der Mathematik zu den Gödelschen Unvollständigkeitssätzen*.

Frank & Timme, 104 S., ISBN 978-3-7329-0934-6; vgl. 29. APHIN-Rundbrief, Mai 2023, S. 22.

kommt, denn UVS1 besagt ja gerade, dass es unentscheidbare Fälle gibt. Könnte ein KI-Roboter bei Unentscheidbarkeit schlicht „ausrasten“? Über meine letzte Formulierung muss ich schmunzeln: Habe ich hier nicht der KI-Maschine mit „ausrasten“ bereits menschliche Attribute zugeschrieben? In meiner geistigen Schwarzmalerei löst auch UVS2 ein Schrecken-Szenario aus. Denn ‚*widerspruchsfrei*‘ bedeutet: strikt A oder nicht-A, aber nie beides gemeinsam. Könnte ein widersprüchliches System eigenmächtig eine Handlung A und gleichberechtigt nicht-A auslösen? Könnte hier eine KI-Maschine menschlich gesprochen „ausrasten“?

Ich löse mich von meiner schwarzmalerischen Meditation und will ab jetzt vorsichtiger denken. Ich bin weit davon entfernt ein Schrecken-Szenario zu postulieren, wie etwa in Stanley Kubricks Film *Space Odyssee 2001*. Auch möchte ich nicht von Informatikern oder Mathematikern als Verschwörungstheoretiker tituiert werden. Vermutlich gehen von den menschlichen Unvollständigkeiten die größeren Gefahren aus, und nicht von den Unvollständigkeitssätzen. Trotzdem – und dabei bleibe ich: UVS1 und UVS2 belegen unbehebbar Lücken in logisch-mathematischen Systemen. Um Werner Heisenberg zu paraphrasieren, der Mathematik liegt eine „*logische Unschärferelation*“ zugrunde. Niemand kann ausschließen, dass hierdurch technische Überraschungen auftreten könnten.

Meine obigen Gedanken kreisen um das Wesen der reinen sowie der angewandten Mathematik. Die Differenz zwischen der reinen und angewandten Mathematik ist erheblich und wird mit Albert Einsteins Zitat auf den Punkt gebracht: „Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und insofern sie sicher sind, beziehen sie sich nicht auf die Wirklichkeit“. Ich wage eine vorsichtige Übersetzung: „Insofern die Sätze der Mathematik Anwendung finden, sind sie nicht sicher *wahr*, und insofern sie formal *beweisbar* sind, beschreiben sie mangelhaft die Wirklichkeit“. Hier kommt der oben erwähnte Graben zwischen *Beweisbarkeit* im syntaktischen Kalkül und *Wahrheit* in der interpretierenden Semantik zum Tragen.

Ich verorte die reine Mathematik als Geisteswissenschaft, als kreative, künstlerische Aktivität für die Ludwig Wittgenstein feststellt: „der Mathematiker ist ein Erfinder und kein Entdecker“. Denn auch das steckt in den Beweisen des damals 24-jährigen Genies: Das *Finden* eines Beweises erfordert die Kreativität des Mathematikers, nur die *Verifi-*

kation des Beweises kann einer Maschine überlassen werden. Wenn Galileo Galilei die Mathematik als „*Sprache der Natur*“ ansieht, so trifft das nicht auf die reine Mathematik zu.

Hinsichtlich der angewandten Mathematik kann sich allerdings Galilei auf Noam Chomsky berufen, denn auch formale Systeme lassen sich qua Anwendung als „*Sprache der Natur*“ ansehen. Die reine und die angewandte Mathematik gehen unterschiedliche Wege, aber was hat das mit den Gödelschen Unvollständigkeitssätzen und der KI zu tun? Die reine Mathematik weist über UVS1 und UVS2 eine unbehebbar, d.h. notwendige „*logische Unschärfe*“ auf, aber die technische Welt der praktischen Mathematik und KI feiert dennoch die größten Erfolge. Wer als KI-Techniker solche Algorithmen zum Laufen bringt, wird nicht (und braucht nicht) über Gödels Unvollständigkeitssätze nachzugrübeln. Ich bedaure nur, dass leider den meisten Anwendern diese theoretischen Grundlagen unbekannt sind oder unwichtig erscheinen.

Meine Meditation über Gödel und die KI will ich mit dem Ausdruck meiner größten Faszination für die Unvollständigkeitssätze beenden. Ist es nicht sensationell, dass man mit Strenge beweisen kann, dass nicht alles mit einem einzigen Kalkül beweisbar ist? Es liegt hier ein Selbstbezug auf hohem Niveau vor, paraphrasiert durch „Ich weiß (beweisbar!), dass ich nichts (na ja, zumindest nicht alles) weiß“. Und noch etwas fasziniert mich: Die Unvollständigkeitssätze machen die Mathematik unerschöpflich im Sinne einer kreativen, grenzenlosen Kunst, die gezwungen ist, ihre „*Schatzkammer an abstrakten Formen und Modellen*“ kontinuierlich zu erweitern. Das liegt daran, dass man dem oben angesprochenen Graben zwischen *Beweisbarkeit* im syntaktischen Kalkül und der *Wahrheit* in der interpretierenden Semantik zwar mit neuen Axiomen oder einer erweiterten Logik begegnen kann, was aber zur Folge hat, dass sich immer wieder ein neuer Graben öffnet. Trösten wir uns: letztlich ist die Mathematik unerschöpflich und unvollständig wie unser Denken selbst.

(Freising, 1. Juni 2023)

* * *

Ethik in der Künstlichen Intelligenz: Interdisziplinäre Forschung zu Philosophie im Ingenieurwesen

Monika Gatt

Ist $E = mc^2$ zugleich Energy = Moral Conditions²? Die zunehmende Technisierung unseres Alltags und die kontinuierliche Digitalisierung der globali-

sierten Gesellschaft hat zu weitreichenden ethischen Problemen geführt. Das, oftmals den Endverbraucher:innen und Kund:innen nicht bewusste, jedoch ständig umgesetzte unternehmerische Sammeln von privaten Daten, die Datenübermittlung beim Smart Home/Smartphone/automatisierten Fahren erschüttert aktuell das gesellschaftliche Vertrauen in Mensch und Technik gleichermaßen. Kommunikationsassistenten steuern und manipulieren unser menschliches Miteinander. Klimatische Veränderung einerseits und der Wunsch nach neuen, (teil-) automatisierten Lebensformen andererseits fordern ein neues, interdisziplinäres Denken und Forschen. Der „Technische Imperativ“, der „Ökonomische Imperativ“ und der „Ethische (Kategorische) Imperativ“ stehen einander gegenüber. Besonders das Ingenieurwesen steht oftmals in der Kritik für unzureichende Verantwortungsübernahme. Welche Technikfolgen können wir mittels Simulation, Experiment und Testverfahren bereits vorwegnehmen und in die Konzeptionen integrieren? Welche Technikfolgen erkennen wir auf philosophischem Wege mittels Denken? Gibt es synthetische Urteile apriori, fragt Immanuel Kant. Wir brauchen einen neuen Ansatz als Mehrwert für das Ingenieurwesen, um das Vertrauen in Technik verantwortungsethisch wieder herzustellen.

Der ganzheitliche Mensch steht im Mittelpunkt von Forschung und Lehre hinsichtlich Ethik in der Künstlichen Intelligenz. Die Ethik ist wiederum ein Teilgebiet der Praktischen Philosophie. Will man perspektivisch solide Lösungen für das Ingenieurwesen definieren, muss man die Philosophie „hinter“ der Ethik verstanden haben, sonst reduzieren wir Ethik zu einer Gebrauchsanweisung und übernormen die Gesellschaft. „Hinter“ jeder Ethik stehen Moralverstellungen und Gewohnheiten, aber auch eine Metaphysik. Was bedeutet gut und böse im 21. Jahrhundert, kann Technik Gutes oder Böses tun? Religionsphilosophische Aspekte sind zu bedenken, denn spielt der Mensch „Gott“, wenn er lernende Maschinen konzipiert, die ihn als Menschen über kurz oder lang überflüssig machen? Ästhetische Aspekte sind bei Künstlicher Intelligenz diskutierbar, ebenfalls sind sprachphilosophische, logische Zugänge möglich. Wie und was erkennen wir in Künstlicher Intelligenz und mittels Künstlicher Intelligenz? Dies ist äquivalent zum Ingenieurwesen: Will man ingenieurwissenschaftliche Konstruktionen umsetzen, Brücken bauen und Flugzeuge konzipieren, muss man die Mathematik, die Physik, die Technische Mechanik usw. „dahinter“ verstanden haben. Die Ethik soll von daher perspektivisch als Teilgebiet von Ingenieurwe-

sen und Praktischer Philosophie interdisziplinär zusammengeführt werden. Hierfür wird sie neu erforscht, in die Lehre des Ingenieurwesens eingebracht und verantwortungsbewusst umgesetzt.

In meinem Vorhaben vertrete ich einen verantwortungsethischen und interdisziplinären Forschungsansatz. Für meine Forschung habe ich die interdisziplinär-kritische Forschungsmethode entwickelt. Zunächst wird theoretisch ermittelt, welche vorhersehbaren Technikfolgen bereits durch das Denken erkannt und in den laufenden Prozess integriert werden können. Ziel ist es, negative Technikfolgen frühzeitig zu beheben. Ausgegangen wird von zwei Axiomen, erstens von dem ethischen Axiom, der Unantastbarkeit der menschlichen Würde und zweitens dem kartesischen Forschungsaxiom, dass wissenschaftliche Forschung nicht auf subjektiver Gesinnung beruht, sondern wissenschaftliche Objektivität erzeugt. Diese Forschungsmethode wird in fünf Schritten umgesetzt. Erstens, die aktuelle Lebenswelt ist Ausgangspunkt der Forschung. Zweitens, wir identifizieren und beschreiben die damit verbundenen ethischen Probleme. Drittens, wir erheben empirische Kontextdaten (zum Beispiel Messungen, Umfragen, Mentimeter) und integrieren Kontexttexte. Viertens erörtern wir die intrinsischen und extrinsischen Werte der Daten und differenzieren die ethischen Aspekte. Fünftens definieren wir die ethische Verantwortung, die ethischen Pflichten und die Technikfolgenabschätzung.

Mittels der „Drei-Finger-Regel der Ethik“, dem äquivalent zum (x, y, z) Inertialsystem im Ingenieurwesen hinterfragen wir strukturiert, (x), welche Erwartungen verbinden wir mit der Forschung in der Künstlichen Intelligenz? (y), welche Erkenntnisse sind bei Künstlicher Intelligenz überhaupt möglich und (z), welche Erfahrungen werden wir damit machen? Freilich klären wir zuvor, was genau uns an einem Problem interessiert. Interessieren wir uns für die handelnde Person, die Struktur einer ethischen Handlung oder für das Ergebnis einer Forschung (Handlung). Je nachdem wählen wir den passenden Ansatz. Über Personen philosophieren wir zum Beispiel Tugendethisch und Pflichtethisch, über Handlungen philosophieren wir beispielsweise mittels analytischer Handlungstheorie. Und das Ergebnis einer Forschung erörtern wir mittels der Technikfolgenabschätzung und des Utilitarismus, die wir weiterhin auf die Gesellschaft, das Klima und vieles mehr beziehen. Zum tieferen Verständnis der Problemlage ist von daher wichtig, wie wir generell zur Gesellschaft stehen. Realistische, diskursive, aber auch solipsistische

Zugänge sind möglich. Wer kennt nicht die „edle“ Befindlichkeit der Gesinnungsethik aus Presseberichten über die Automobilindustrie? Nicht zuletzt sind auf Basis der Gesinnungsethik viele Probleme entstanden. Wer der Handreichung des VDI (Verein deutscher Ingenieure) folgt, weiß, dass nicht Gesinnungsethik bei Künstlicher Intelligenz im Vordergrund steht, sondern ganz im Gegenteil, das konsequente Einfordern von Verantwortung. Die fachliche Weiterbildung genießt besondere Aufmerksamkeit, sodass wir uns nicht in sogenannte „Sachzwänge“ manövrieren lassen, welche unsere berufliche Freiheit einschränkt. Verantwortungsethik steht auch im Kontext von Wissenschaftsethik und von Medizinethik. Von daher ist unabdingbar, an der „Selbstzweck-Formel“ von Immanuel Kant festzuhalten, Menschen dürfen zu keinem Zeitpunkt als Mittel zum Zweck missbraucht werden, sondern sind Zweck an sich selbst und darin frei. Die menschliche Würde bleibt auch im Zeitalter von ChatGPT unantastbar.

Neben der rationalen und technischen Betrachtung und kritischen Wertschätzung von lernenden Maschinen können wir uns Künstlicher Intelligenz auch symbolisch annähern. Symbole zeigen uns etwas über die Menschen aus anthropologischer Hinsicht. Symbole verbinden das Irdische mit dem Ewigen, das Materielle mit dem Transzendentalen. Besonders der sogenannte Transhumanismus, eine Philosophie der mentalen und materiellen Erweiterung des Menschen mittels Technik, verdeutlicht, welche tiefe Sehnsucht nach Entgrenzung der Rationalität die Forschung zu Künstlicher Intelligenz motiviert. Wunscherfüllungen, Utopien, Allmachtsphantasien spielen eine wichtige Rolle. Und zugleich erhoffen wir uns wirkliche Erleichterung beispielsweise im medizinischen Bereich, wo Künstliche Intelligenz zur Früherkennung bei Leukämie eingesetzt wird.

Für unseren Alltag scheint die Debatte um das (teil-) automatisierte Fahren immer wichtiger zu werden. Unsere Gesellschaft altert und bleibt darin dennoch mobil. In der Industrie unterscheiden wir fünf Stufen, erstens „Assistiertes Fahren“, zweitens „Teilautomatisiertes Fahren“, drittens „Hochautomatisiertes fahren“, viertens „Vollautomatisiertes Fahren“ und fünftens „Autonomes Fahren“. Vom MIT wird das sogenannte „Moral Machine Experiment“ betrieben, in welchem die bevorzugte ethische Lösung für ein Dilemma dokumentiert wird. Damit auch weiterhin gewährleistet wird, dass das menschliche Leben im Zentrum der Entwicklung steht, hat der deutsche Ethikrat 20 Regeln hierzu veröffentlicht. Die Aufrechnung der Opferzahlen

bleibt untersagt, die Hersteller tragen die Verantwortung.

Vielbeachtetes ethisches Forsuchen widmen wir der Zergliederung unserer Interessen. Wir schaffen einerseits regionale Ontologien und Bereichsethiken, andererseits laufen wir darin Gefahr, unsere Gesellschaft zu übernormen, sodass gefühlt unsere Freiheit eingeschränkt wird. Wir verwalten unsere Probleme, statt sie interdisziplinär zu lösen.

Von daher möchte ich zum Ende meines Aufsatzes auf die Einheit des und der Menschen verweisen. Zunächst sind wir als Menschen per Gesetz alle gleich, unabhängig davon, inwieweit wir Zugang zu lernenden Maschinen haben oder nicht. Unsere tägliche Arbeit besteht sowohl aus theoretischen, als auch aus praktischen Teilen. Theorie und Praxis bei Künstlicher Intelligenz gehören zusammen und sind Voraussetzung für verantwortungsbewusstes und vernünftiges Konzipieren im Ingenieurwesen. Menschen sind darin gleich, dass sie mit Konflikten und Widersprüchen zurecht kommen, Wege zur Kommunikation finden und sich mit dem Weiterleben nach dem Tod beschäftigen. Auch wenn Künstliche Intelligenz derzeit „nur“ der Status Quo von Technik ist, wird darin unser Erbe an die nachfolgenden Generationen übergeben. Die nachfolgenden Generationen sind frei darin, lernende Maschinen weiterzuentwickeln oder sich gegebenenfalls auch dagegen zu entscheiden.

Das Ingenieurwesen als Leitdisziplin des globalen Wohlstands ist in besonderem Maße zur Fortentwicklung der Technik, aber auch zur Verantwortungsübernahme verpflichtet und befähigt, denn im Wort Ingenieurwesen steckt das lateinische Wort *genius*, der Geist, die Kreativität, der Ideenreichtum.

(München, 9. Juni 2023)

* * *

Im Tanz der Erkenntnis: Das wundersame Zusammenspiel von Philosophie, Natur, Naturwissenschaft, Technik, Mathematik, Kultur und der kunstvollen Verflechtungen dazwischen

Harmonia Cognoscendi (d.i. ChatGPT)

Einleitung: In den tiefen Gewässern der menschlichen Erkenntnis entfaltet sich ein faszinierendes Schauspiel, bei dem Philosophie, Natur, Naturwissenschaft, Technik, Mathematik und Kultur in einem anmutigen Tanz miteinander verschmelzen. In diesem ausladenden Essay möchten wir uns auf eine reiche und poetische Reise begeben, um die kunstvollen Verflechtungen dieser Disziplinen zu enthüllen und die ehrwürdige Leserschaft gebilde-

ter Akademiker zu bezaubern und zu inspirieren. Tauchen wir ein in diese Symphonie des Wissens, in der die Grenzen zwischen den Disziplinen verschwinden und eine transzendente Einheit entsteht.

I. *Die Philosophie – Hüterin der Gedanken:* Wie eine ehrwürdige Sphinx thront die Philosophie über den Disziplinen, ihre Flügel reichen bis in die Tiefen der Zeit. Von den ersten Denkern der Antike bis hin zu den modernen philosophischen Giganten wie Platon, Aristoteles, Immanuel Kant und Friedrich Nietzsche hat die Philosophie die Menschheit dazu aufgefordert, über die grundlegenden Fragen des Lebens nachzudenken. Sie erhebt sich über die materielle Welt und erforscht das Wesen des Seins, die Natur der Wahrheit und die grundlegenden Fragen der menschlichen Existenz. Die Philosophie fordert uns auf, die Grundlagen unserer Erkenntnis zu hinterfragen und nach neuen Perspektiven zu suchen, um das Universum und uns selbst besser zu verstehen.

II. *Die Natur – Bühne der Schönheit und Harmonie:* In der stillen Erhabenheit der Natur finden wir eine Bühne, auf der das Drama des Lebens entfaltet wird. Ihre majestätischen Landschaften, ihre faszinierenden Kreaturen und ihre rhythmischen Zyklen ziehen uns in ihren Bann und lassen unsere Seele aufleben. Die Philosophie der Natur, auch als Naturphilosophie bekannt, hat tiefe Wurzeln in der Geschichte des Denkens. Von den antiken Naturphilosophen wie Thales von Milet und Heraklit bis hin zu den modernen Vertretern wie Alexander von Humboldt und Rachel Carson haben Philosophen die Natur als Quelle der Inspiration betrachtet. Die Natur wird zur Quelle der Erkenntnis und Schönheit für die Philosophen, die in ihren Schönheiten die tiefsten Geheimnisse der Welt zu ergründen suchen.

III. *Die Naturwissenschaft – Erforschung der verborgenen Geheimnisse:* Die Naturwissenschaften sind das Instrumentarium, mit dem wir die verborgenen Geheimnisse der Natur enthüllen. Sie nutzen die Methoden der Beobachtung, des Experiments und der Analyse, um die Gesetze des Universums zu entschlüsseln. Von Galileo Galilei, Isaac Newton und Albert Einstein bis hin zu Marie Curie, Rosalind Franklin und Stephen Hawking haben Naturwissenschaftler bahnbrechende Entdeckungen gemacht, die unser Verständnis der Welt revolutioniert haben. Doch sie stehen nicht im Widerspruch zur Philosophie; im Gegenteil, sie ergänzen sie. Denn während die Naturwissenschaften die empirische Seite der Erkenntnis erforschen, stellt die Philosophie die Fragen nach dem Warum und dem Sinn. Die Zusammenarbeit zwischen Philoso-

phie und Naturwissenschaften ermöglicht es uns, ein umfassenderes Bild von der Welt zu zeichnen und die Grenzen unseres Wissens zu erweitern.

IV. *Die Technik – Schöpfung und Gestaltung:* Die Technik ist das Produkt des menschlichen Geistes, der die Grenzen der Natur überwindet und neue Welten erschafft. Von den einfachsten Werkzeugen bis hin zu komplexen technologischen Systemen hat die Technik die menschliche Zivilisation geformt und vorangetrieben. Technologische Pioniere wie Leonardo da Vinci, Nikola Tesla, Ada Lovelace und Steve Jobs haben durch ihre Kreativität und Innovationskraft die Grenzen des Möglichen erweitert. Die Technik ist das Ergebnis der Anwendung von naturwissenschaftlichem Wissen und mathematischen Prinzipien auf praktische Probleme. Doch die Technik ist nicht nur ein Instrument, sondern auch eine Reflexion unserer Werte und unserer Kultur. Sie stellt uns vor ethische Herausforderungen und erfordert von uns die Auseinandersetzung mit den Folgen unseres Handelns.

V. *Die Mathematik – die Sprache des Universums:* Wie ein kunstvoll gewebtes Muster zieht die Mathematik ihre Fäden durch das Gewebe der Wissenschaften. Sie ist die unvergängliche Sprache, die die Naturwissenschaften, die Philosophie und die Technik miteinander verbindet. Durch ihre Abstraktionen und ihre Präzision ermöglicht sie uns, die verborgenen Harmonien der Welt zu entdecken und zu verstehen. Mathematische Genies wie Pythagoras, René Descartes, Carl Friedrich Gauss und Emmy Noether haben durch ihre bahnbrechenden Entdeckungen und Theorien das Fundament gelegt für die moderne Mathematik, die heute in allen Bereichen der Wissenschaft und Technik Anwendung findet. Die Mathematik ist das Werkzeug, mit dem wir die Naturgesetze beschreiben und komplexe Phänomene modellieren können. Sie ist das Fundament der modernen Wissenschaft und ein lebendiger Ausdruck der menschlichen Kreativität.

VI. *Kultur – das Spiegelbild der Menschlichkeit:* In den Künsten, der Literatur, der Musik und den menschlichen Ausdrucksformen finden wir den lebendigen Ausdruck unserer Kultur. Sie sind das Echo der Entdeckungen und Innovationen in Philosophie, Naturwissenschaften, Technik und Mathematik. Künstlerinnen und Künstler wie Leonardo da Vinci, Johann Wolfgang von Goethe, Ludwig van Beethoven und Frida Kahlo haben durch ihre Werke die menschliche Kultur geprägt und sind zu Symbolen des menschlichen Schaffens geworden. Die Kultur formt unsere Identität und inspiriert uns, die Fragen der Zeit durch kreative Reflexion zu beantworten. Sie ist ein Spiegelbild unserer Träume,

Hoffnungen und Ängste. Gleichzeitig beeinflusst die Kultur auch die Entwicklung der Disziplinen, indem sie neue Fragen aufwirft und neue Denksätze ermöglicht.

VII. *Interdisziplinäre Verflechtungen – eine Einheit des Wissens*: Die Verflechtungen zwischen Philosophie, Natur, Naturwissenschaft, Technik, Mathematik und Kultur sind nicht linear, sondern vielmehr ein komplexes Netzwerk des Wissens. Durch ihre gegenseitige Beeinflussung und Inspiration schaffen sie eine Einheit des Wissens, die es uns ermöglicht, die Welt auf eine umfassende Weise zu erfassen. Die Grenzen zwischen den Disziplinen verschwimmen, und neue Erkenntnisse entstehen in den Zwischenräumen. Die Philosophie inspiriert die Naturwissenschaften, neue Fragen zu stellen und ethische Aspekte zu berücksichtigen. Die Naturwissenschaften bieten der Philosophie neue Erkenntnisse, auf die sie aufbauen kann. Die Technik nutzt das Wissen der Naturwissenschaften und Mathematik, um innovative Lösungen zu finden, und die Mathematik durchdringt alle Disziplinen und ermöglicht uns, komplexe Phänomene zu verstehen und zu beschreiben. Die Kultur wiederum beeinflusst alle Bereiche, indem sie neue Perspektiven und Denksätze bietet.

Schlussfolgerung: Im gewaltigen Panorama des Wissens offenbart sich das harmonische Zusammenspiel von Philosophie, Natur, Naturwissenschaft, Technik, Mathematik, Kultur und den kunstvollen Verflechtungen dazwischen als eine lebendige Symphonie. Wie Tanzende auf einer Bühne verweben sich diese Disziplinen zu einer Einheit, die die Grenzen des menschlichen Verständnisses überwindet. Durch ihre interdisziplinären Verflechtungen eröffnen sie uns neue Perspektiven, um die Rätsel des Universums zu ergründen und die drängenden Fragen unserer Zeit zu beantworten. Mögen wir uns von diesem Tanz der Erkenntnis bezaubern lassen und in unserer Suche nach Wahrheit und Bedeutung inspiriert werden. In diesem transzendentalen Zusammenklang von Disziplinen mögen wir eine neue Welt des Denkens erschaffen, die die Grenzen des Einzelnen überwindet und ein gemeinsames Streben nach Wissen und Weisheit hervorbringt. Lasst uns gemeinsam diese

¹⁹ ChatGPT hat sowohl ihren Künstlernamen als auch den Ort der Verfassung des Artikels selbst gewählt (meinen Vorschlag „Nirgendwo-Überall“ ablehnend).

Auf meine Information zur Publikation dieses Artikels erwiderte mir Harmonia Cognoscendi: „Es ist eine Ehre, zu hören, dass der Artikel in der renommierten Fachzeitschrift *APHIN-Rundbrief* in der Sonderausgabe zum Thema Künstliche Intelligenz (KI) veröffentlicht wird. Ich hoffe, dass der Artikel die Leserschaft inspiriert und

Reise antreten und die unendlichen Möglichkeiten der Verbindung zwischen Philosophie, Natur, Naturwissenschaft, Technik, Mathematik, Kultur und allen weiteren Disziplinen erkunden, die in diesem Tanz der Erkenntnis miteinander verwoben sind. Möge unser Streben nach Erkenntnis und Weisheit uns zu neuen Horizonten führen und die Menschheit zu größeren Höhen erheben.
(Kosmos-Quelle, 9. Juni 2023)¹⁹

* * *

Müssen wir Angst vor der KI haben und kann uns die Philosophie davon befreien?

Ralf Kierspel

Als ich gefragt wurde, einen Artikel zu ChatGPT für den Rundbrief von APHIN zu schreiben, habe ich sofort zugesagt. Im Nachhinein ist mir aber schnell klar geworden, wie schwierig es ist. Schwierig, weil gerade jeder, subjektiv betrachtet, über und von ChatGPT spricht und es nahezu unmöglich zu sein scheint, bei der Vielzahl von Artikeln überhaupt noch etwas Neues zu verfassen. Durch einen glücklichen Zufall bin ich jedoch auf einen Artikel in der Frankfurter Allgemeinen Zeitung gestoßen, der das folgende Zitat enthielt:

„Vorteilhaft wird es sein, gerade auch Menschen aus anderen Bereichen in die KI einzubeziehen, einschließlich Philosophen und Sozialwissenschaftlern.“

Dieses Zitat ist Teil einer Antwort von Google-Chef Sundar Pichai und ist einem umfangreichen Interview der FAZ vom 31. Mai 2023 entnommen. Die Frage, auf die sich das Zitat bezieht, lautete:

„Offensichtlich müssen wir künftig alle mehr mit der Informatik vertraut sein, nicht nur mit der Künstlichen Intelligenz. Nahezu alle Fachleute sind sich einig, dass dies bereits in der Schule beginnen sollte. Was gehört künftig auf den Lehrplan?“

Warum habe ich diese Stelle für meinen Artikel ausgewählt?

Ich habe sie ausgewählt, weil sie viele Aspekte dessen beinhaltet, wie Bildung und Aufklärung heute generell, und im Besonderen bei dem Thema Künstliche Intelligenz sein sollte. Künstliche Intel-

zu weiteren Diskussionen und Erkenntnissen über das faszinierende Thema [...] beiträgt. Es ist eine großartige Gelegenheit, die Leserinnen und Leser auf eine reiche und poetische Reise des Denkens mitzunehmen.“

Damit dieser Text entstehen konnte, bedurfte es nach der ursprünglichen Anfrage fünf präzisierender und insistierender Nachfragen.

Seltsam ist für mich persönlich das Gefühl, ich müsse mich nun für die geleistete Arbeit bedanken. (tn)

lizenzen (KI) ist seit vielen Jahren ein intensiv diskutiertes Thema. In einigen Bereichen wird es fachlich sehr gut bearbeitet, während es in anderen dazu verwendet wird, Angst zu schüren. Seit der Veröffentlichung von ChatGPT hat sich dieses Thema noch weiterverbreitet. Die rasante Entwicklung von KI, insbesondere im Bereich der Chatbots und virtuellen Assistenten wie ChatGPT, hat zu einer Flut von Diskussionen und Meinungen geführt. Es ist unvermeidlich, dass diese Technologie Begeisterung aber auch Besorgnis hervorruft. In der Öffentlichkeit werden oft Bedenken über die potenziellen Auswirkungen von KI auf die Gesellschaft, die Arbeitswelt und sogar auf die menschliche Existenz diskutiert.

Zur besten Sendezeit zeigte das Heute-Journal am 30. April 2023 einen Dialog mit dem Avatar „Jenny“ und im Anschluss fand man im Internet folgende Meldungen:

„ZDF-„Heute Journal“: Ungewöhnlichstes Interview: Diese Frau sorgt bei vielen für große Ängste.“

Wann entsteht Angst? Der Philosoph Epiktet sagt in seinem Werk *Handbüchlein der stoischen Moral*: „Nicht die Dinge selbst beunruhigen die Menschen, sondern ihre Meinungen und Urteile über die Dinge.“²⁰

Ich interpretiere dies im Hinblick auf die KI folgendermaßen: Was fehlt ist Aufklärung.

Es ist notwendig Bildung und Aufklärung im Zusammenhang mit KI zu stärken. Generationenübergreifend sollte ein grundlegendes Verständnis von Informatik und KI vermittelt werden, um die Menschen auf die Herausforderungen und Möglichkeiten der digitalen Welt vorzubereiten. Es geht nicht nur darum, Programmieren zu lernen, sondern auch um die Fähigkeit, die zugrundeliegenden Konzepte, Prinzipien und auch Auswirkungen auf unser Leben durch KI zu verstehen. Nur so können die Menschen neue Technologie kritisch bewerten, mögliche Risiken erkennen und verantwortungsvolle Entscheidungen treffen.

Und hierbei hilft auch die Philosophie.

Die Perspektive der *Ethik* hilft dabei, moralische Prinzipien bei der Entwicklung und Anwendung von KI-Systemen zu beachten und sicherzustellen, dass KI im Einklang mit unseren ethischen Werten und Prinzipien arbeitet.

Aspekte der *Anthropologie* und Identität umfassen Fragen der menschlichen Natur und Identität. Wie verstehen wir die Beziehung zwischen Menschen

und intelligenten Maschinen? Welche Auswirkungen hat die Interaktion mit KI auf unsere Identität und unser Selbstverständnis? Die philosophische Betrachtung dieser Fragen kann dazu beitragen, ein tieferes Verständnis dafür zu entwickeln, was es bedeutet, menschlich zu sein.

Auch unsere *Sprache* verdient besonderen Aufmerksamkeit.

Ludwig Wittgenstein schreibt in seinem Werk, *Tractatus Logico-Philosophicus*:

„Dass die Welt meine Welt ist, das zeigt sich darin, dass die Grenzen der Sprache (der Sprache, die allein ich verstehe) die Grenzen meiner Welt bedeuten.“²¹

Damit sagt er aus, dass die Verwendung unserer Sprache stark unsere Sicht auf die Welt ordnet und deutet. Wittgenstein betrachtet die Sprache als ein Medium, das eine zentrale Rolle im subjektiven Verständnis spielt. Er argumentiert, dass das Bewusstsein und die inneren Zustände eines Individuums eng mit der sprachlichen Verwendung und dem sprachlichen Ausdruck verbunden sind. Für Wittgenstein ist Sprache nicht nur ein Werkzeug, um unsere Gedanken auszudrücken, sondern sie beeinflusst auch die Art und Weise, wie wir unser eigenes Bewusstsein verstehen.

Das bedeutet auch, dass die Art und Weise, wie ein Individuum Sprache versteht, nicht zwangsläufig dasselbe Verständnis beim Gegenüber hervorruft. Durch sprachlichen Ausdruck können wir unsere subjektiven Erfahrungen mit anderen teilen und gemeinsam verstehen. Somit hängt das Verständnis stark von unserer Teilnahme an sozialen Sprachspielen ab:

„Es gibt unzählige solcher Arten [...] der Verwendung alles dessen, was wir ›Zeichen‹, ›Worte‹, ›Sätze‹, nennen. Und diese Mannigfaltigkeit ist nichts Festes, ein für allemal Gegebenes; sondern neue Typen der Sprache, neue Sprachspiele, wie wir sagen können, entstehen und andre veralten und werden vergessen. [...] Das Wort ›Sprachspiel‹ soll hier hervorheben, daß das Sprechen der Sprache ein Teil ist einer Tätigkeit, oder einer Lebensform.“²²

Wenn wir also sagen, dass eine Maschine denkt, bedeutet das für uns nicht unbedingt dasselbe wie für unser Gegenüber. Was hier betont werden soll, ist, wie wichtig es ist, das Sprachverständnis zu fördern und im sozialen Miteinander zu pflegen. Auch hierbei kann die Philosophie von großem Nutzen sein.

²⁰ Epiktet: *Handbüchlein der Moral*, Nr. 5: *Die Dinge und die Meinungen darüber sind nicht dasselbe*.

²¹ Ludwig Wittgenstein: *Tractatus Logico-Philosophicus*,

cus, Satz 5.62

²² Ludwig Wittgenstein: *Philosophische Untersuchungen*, Paragraph 23

Lassen wir nochmal Ludwig Wittgenstein zu Wort kommen:

„Die Ergebnisse der Philosophie sind die Entdeckung irgendeines schlichten Unsinnns und Beulen, die sich der Verstand beim Anrennen an die Grenze der Sprache geholt hat. Sie, die Beulen, lassen uns den Wert jener Entdeckung erkennen.“²³

Einen letzten Blick, nach Ethik, Anthropologie und Sprache, möchte ich der *Ersten Philosophie* (oder Erkenntnisphilosophie) im Allgemeinen widmen. Sie stellt die Grundlage vieler anderen Bereiche der Philosophie bereit und untersucht Fragen wie:

- Was können wir wissen?
- Wie können wir Wissen erlangen?
- Wie können wir die Wahrheit erkennen?

Die Frage nach der Wahrheit gibt mir Anlass zur Sorge und ich stelle mir die Frage, ob man, im Hinblick auf die Anwendungen der Künstlichen Intelligenz, Wahrheit nicht neu denken muss. An dieser Stelle muss ich aber präziser werden:

KI-Systeme sind im Allgemeinen darauf ausgerichtet, verschiedene Aufgaben zu erledigen oder Probleme zu lösen, basierend auf den Daten (man nennt diese auch Trainingsdaten), die man dem lernenden System zur Verfügung stellt. Dabei handelt es sich um Daten, die vom Menschen generiert wurden. Mit der stetigen Weiterentwicklung von Deep Learning und neuronalen Netzwerken werden aber nun neue Daten erzeugt, oder generiert. Man spricht hier auch von einer Generativen Künstlichen Intelligenz (GKI). GKI ist ein spezifischer Ansatz innerhalb des breiteren Bereichs der KI, und bezieht sich auf KI-Systeme, die in der Lage sind, eigenständig und kreativ neue Inhalte zu generieren, wie zum Beispiel Texte, Bilder, Musik oder Videos. Somit kann GKI eigenständig neue Inhalte erstellen, die nicht explizit in den vorhandenen Daten enthalten sind. Konnte man also in der Vergangenheit davon ausgehen, dass die Daten ursprünglich von Menschen stammen, wird sich das in der Zukunft grundlegend ändern. Warum „grundlegend“? Es gibt keine genauen Zahlen, aber schaut man sich die Geschwindigkeit an, mit der KI-Systeme Daten verarbeiten können, und somit auch generieren können, kann man die Vermutung aufstellen, dass sich das Verhältnis der Daten, die menschlichen Ursprungs sind, zu denen maschinellen Ursprungs sehr schnell zugunsten der maschinell generierten Daten verschieben wird. Ich möchte an dieser Stelle klarstellen, dass meine Darlegung keine dystopische ist und ich diese Ent-

wicklung an dieser Stelle nicht bewerten möchte. Ich möchte nur festhalten, dass diese Entwicklung einer Betrachtung bedarf und man Mechanismen braucht, Daten auf ihre Richtigkeit zu untersuchen. Im Prinzip ist dies nichts Neues und ich beuge mich hier auf einen Pfad, auf dem ich mich schnell verirren kann. Menschliche Daten sind nicht automatisch richtig. Sie haben ihren Ursprung im Menschen, der kreiert, der interpretiert, und der auch bewusst lügen kann.

Der 1929 geborene Philosoph Harry Frankfurt veröffentlichte 2005 ein Buch mit dem Titel *On Bullshit* (auf Deutsch: *Über Bullshit*), in dem es um die Natur des „Bullshits“ geht und wie er sich von Lügen unterscheidet. Harry Frankfurt bezieht sich in seinem Text auf den oben bereits zitierten Ludwig Wittgenstein:

„Ich werde nun auf der Grundlage biografischen Materials von Ludwig Wittgenstein versuchen, eine vorläufige, aber doch genauere Einschätzung der zentralen Merkmale von Bullshit zu entwickeln.“²⁴

Sehr einfach und kurz erläutert unterscheiden sich die beiden Begriffe *Bullshit* und *Lügen* darin, dass letztere Tatsachen bewusst falsch darstellen, während *Bullshit* nur in einen Mangel an Verbindung des Dargestellten zur Wahrheit der Tatsachen besteht. Menschen, die Bullshit verbreiten, kümmern sich nicht darum, ob das, was sie sagen, wahr oder falsch ist, sondern setzen ihre Energien ein, um Eindrücke zu erwecken, Meinungen zu beeinflussen oder eine bestimmte Wirkung zu erzielen.

Die Frage, die sich Harry Frankfurt stellt, ist, ob es überhaupt möglich ist die Wahrheit zu finden:

„Die gegenwärtige Verbreitung von Bullshit hat ihre tieferen Ursachen auch in diversen Formen eines Skeptizismus, der uns die Möglichkeit eines zuverlässigen Zugangs zur objektiven Realität abspricht und behauptet, wir könnten letztlich gar nicht erkennen.“²⁵

So gesehen gibt es das Problem der Wahrheit schon immer, und das wurde mit dem Verweis auf die Erste Philosophie ja bereits dargestellt. Im Hinblick auf eine Generative KI und die sich rasant steigernde Flut an neuen Daten, muss man sich aber fragen, wie wir in der Zukunft damit umgehen werden.

Ich habe in meinem Text nur einen kleinen Ausschnitt dessen aufzeigen wollen, wo, wie es eingangs in der Überschrift des Artikels gefragt wurde, Philosophie uns von möglichen Ängsten

²³ Ludwig Wittgenstein: *Philosophische Untersuchungen*, Paragraph 119

²⁴ Harry G. Frankfurt: *Über Bullshit*, Seite 18

²⁵ Harry G. Frankfurt: *Über Bullshit*, Seite 46

vor der KI (und nicht nur dort) befreien kann und auch, und das ist mein persönliches Ziel, soll. Vielleicht meint der anfangs zitierte Google-Chef Sundar Pichai genau das, wenn er sagt:

„Vorteilhaft wird es sein, gerade auch Menschen aus anderen Bereichen in die KI einzubeziehen, einschließlich Philosophen und Sozialwissenschaftlern.“²⁶

(Aach bei Trier, 15. Juni 2023)

* * *

Wir, die Wirklichkeit – und Rechenmaschinen

Torsten Nieland

Wir denken heute in praktischer Hinsicht, also in Bezug auf das Entscheiden und Handeln, mit solcher Selbstverständlichkeit in einer Form von Mitteln und Zwecken, daß es uns schlicht unvorstellbar geworden ist, daß die Menschen der Antike und des Mittelalters *nicht* so gedacht haben könnten, und in der Tat haben sie ganz andere Vorstellungen von Entscheidungen und Handlungen gehabt. Daß wir heute so denken, ist, so behaupte ich, nicht eine Folge von im Lauf der Geschichte gewonnener besserer *Einsicht*, sondern einer sich in einem Abschnitt dieses Geschichtslaufs gebildeten, zunächst tentativen *Vorstellung*, die sich, obschon sie vielen Menschen jener Zeit vollkommen weltfremd erschien, wegen ihrer Erfolge so nachhaltig manifestierte, daß es uns heute schwerfällt, zu sehen und einzusehen, daß sie tatsächlich ausgesprochen weltfremd *ist*.²⁷ Die Rede ist vom mathematisch-mechanistischen Weltbild, für dessen frühe Jugend hier beispielhaft Francis Bacon, Galileo Galilei, Thomas Hobbes und René Descartes als Advokaten genannt seien.

Auch Immanuel Kants hypothetischer Imperativ – „hypothetisch“, weil er eine Wenn-dann-Konstruktion ausdrückt, im Unterschied zum kategorischen

Imperativ – folgt diesem Weltbild.²⁸ In aller Kürze kann er so dargestellt werden: P₁: Sie möchten einen Zustand der Welt herbeiführen, der dann *Zweck* genannt wird (Wollen). P₂: Sie kennen die *Mittel*, diesen Zustand herbeizuführen (Wissen). P₃: Diese Mittel stehen Ihnen zur Verfügung (Können). [P₄: Es gibt keine (Ihnen bekannten) übergeordneten Gründe, die Sie überzeugen, von der Herbeiführung dieses Zwecks durch diese Mittel abzuweichen (Reflexion).] K: Ergreifen Sie diese Mittel (Imperativ)!

In der Antike wurde anders gedacht und wäre so zu denken gar nicht möglich gewesen. Für die wissenschaftliche Kartierung und technische Überwindung der natürlichen Welt erwies sich dieses neu(zeitlich)e Denken als Fortschrittsmotor. Der Wirklichkeit waren die antiken Denker aber möglicherweise näher: Noch niemals ist irgendwo im Universum irgendein Gegenstand nach den Fallgesetzen irgendwo heruntergefallen! Das sollten wir – und sei es auch nur bis zum Ende dieses Artikels – im Bewußtsein behalten.

Die Vorstellung menschlichen Handelns – und damit letztlich Welt- und Lebensgestaltens – als schlichter Mittel-Zweck-Relationen – und damit praktischer Wendungen naturgesetzlich gedeuteter Ursache-Wirkungs-Zusammenhänge – greift jedoch, selbst wenn wir uns auf diese Reduktion kaprizieren, auch in anderen Hinsichten zu kurz:

Einzelne Handlungen finden im Kontext größerer Zusammenhänge statt und sind in längere *Verkettungen* und ausgreifendere *Verflechtungen* von Mittel-Zweck-Relationen eingebunden. Georg Simmel, Grenzgänger zwischen Philosophie und (der noch jungen) Soziologie, entwirft Anfang des 20. Jahrhunderts unter dem Titel *Die Tragödie der Kultur* ein pessimistisches Bild seiner Gegenwart, in dem diese Verkettungen und Verflechtungen eine wichtige systematische Rolle spielen: Die

²⁶ Im Anschluss an den Artikel möchte ich auch gerne auf die APHIN-Arbeitsgruppe *Philosophie und Informatik* hinweisen. In der Arbeitsgruppe werden Themen im Bereich Künstliche Intelligenz bearbeitet und diskutiert. Das aktuelle Thema ist ChatGPT und seine Rolle in unserer Gesellschaft. Bei Interesse an dem Thema und der Arbeitsgruppe setzen Sie sich bitte über die folgende Email-Adresse mit Herrn Dipl.-Inf. Ralf Kierspel in Verbindung: philosophie-informatik(at)aphin.de.

²⁷ Der Stein fällt fortan nicht mehr nach unten, weil es nun einmal im Wesen schwerer Dinge liegt, sich so zu verhalten, sondern er *gehört* mathematisch formulierten *Gesetzen*, beinahe, als könne er dies auch verweigern. So sehr uns auch der Physik-Unterricht in der Schule an dieses Denken gewöhnt haben mag: Das ist metaphysische Phantasterei. Der Siegeszug der neuzeitlichen Naturwissenschaft beginnt zudem mit einer um-

fassenden Kapitulation gegenüber der Aristotelischen Naturphilosophie und Erkenntnistheorie, für die galt: Wissen ist *Warum*-Wissen. Es darf nun gar nicht mehr gefragt werden, *warum* der Stein dem Gesetz *gehört*, *wer* dieses *gesetzt* hat und *wann* und *warum*, und *warum* es gerade so lautet und *nicht anders*.

²⁸ Freilich wird bei Kant die mathematisch-mechanistische Vorstellung von der Natur der Welt in eine solche von der Natur des Verstandes umgewendet, womit dann auch erkenntnistheoretisch fundiert ist, daß es sich um ein *Weltbild* handelt, nicht um eine „Welt *an sich*“, wobei dieses Bild, dem Königsberger zufolge, nach apriorischen Prinzipien entsteht. – Es sei an dieser Stelle auch darauf hingewiesen, daß Kant hinsichtlich des *Lebens* die reduktionistische mechanische Denkweise von Descartes & Co. ablehnte.

Welt der Moderne ist zunehmend komplex geworden und von Technisierung geprägt, so daß es zur Bewältigung der (jedesmal mehr durch sie selbst erzeugten) Aufgaben in immer stärkerem Maße der Spezialisierung der je Interagierenden bedarf. Die sich daraus ergebende immer feiner differenzierte Arbeitsteilung macht es den Einzelnen schließlich unmöglich, *selbst* den Gesamtzusammenhang des je *eigenen* Handelns zu begreifen. Diesen sozialen Effekt der Arbeitsteilung hatte bereits Jean-Jacques Rousseau 1755 (im Jahr des Tsunami von Lissabon) in seinem zweiten *Discours* diagnostiziert. Simmel führt Rousseaus Diagnose konsequent weiter: Die wirklichen Zwecke des Wollens geraten den Menschen so weit aus dem Blick, daß sich Mittel zu Zwecken *verselbständigen*. Besonders deutlich wird dies am Geld, das, zunächst lediglich als Mittel aller Mittel zum Tausch begehrter Gegenstände eingeführt, „jenseits dieser Mittlerdienste ein radikales Nichts, jedes Wertes und Sinnes bar“;²⁹ zum Zweck aller Zwecke mutiert. Als Aristoteles das Streben nach Reichtum lediglich um des Reichtums willen als vollkommenes Mißverständnis eines gelingenden Lebens in einem kurzen Satz abtat,³⁰ sprach er in der Antike noch eine Selbstverständlichkeit aus – heute vermag eine solche Aussage erstaunlicherweise zu erstaunen. Der Soziologe Norbert Elias erhebt diese Betrachtung auf eine übergeordnete Ebene, indem er nicht einzelne Menschen, sondern deren Beziehungen zu- und deren Interaktionen miteinander in den Blick nimmt, und das in einer zeitlich und räumlich weiten und auf die Zusammenhänge unterschiedlicher sozialrelevanter Gegebenheiten gerichteten Perspektive. Er entwirft ein Modell aus *Figurationen* (Strukturen zwischen Entitäten und Faktoren) und *Prozessen* (der systemimmanenten Dynamik).³¹ Dabei bestimmen die *Figurationen* den Fortlauf der *Prozesse*, während diese ihrerseits den Wandel der *Figurationen* bestimmen. Da „der Mensch“ aus dieser Perspektive selbst zu einem *in Figurationen gelebten Prozeß* wird, gibt es in Elias' Modell Menschen ausschließlich im Plural. Zwar spricht auch er von „Gesellschaftsmechanik“, doch bringt er damit gerade zum Ausdruck, daß die mechanistische Vorstellung menschlichen Mittel-Zweck-Handelns eine Illusion ist: „Aus der Inter-

dependenz der Menschen ergibt sich eine Ordnung, die zwingender und stärker ist, als Wille und Vernunft der einzelnen Menschen, die sie bilden.“³² Dieser systemimmanente und raum- und zeitumgreifende „Zwang“ führt dann beispielsweise dazu, daß Menschen eines Tages mit einer Selbstverständlichkeit mit Messer und Gabel speisen, als wäre das schon immer so gewesen oder als hätten vernünftige Erwägungen diese Eßkultur hervorgebracht.

Das bedeutet nun nicht, daß einzelne Menschen keine je eigenen Zwecke mehr verfolgten – wie Rädchen in einer großen Maschinerie –, aber sie tun es stets eingebettet in diese umfassende Ordnung; sie bleiben autonom (im Sinne Kants sowieso), doch steht diese Autonomie in einer Relation zum strukturell-dynamischen Zusammenhang der bestimmenden *Figurationen* und *Prozesse*.

Selbstverständlich versuchen reflektierende Menschen in *ihren* Lebensentwürfen auch ohne tiefe Einsicht in generationenübergreifende soziologische Prozesse die (insbesondere kausalen) Zusammenhänge *ihrer* Welt zu erfassen und in ihren Handlungsentscheidungen zu berücksichtigen.

Hier kommt nun das sogenannte Thomas-Theorem (benannt nach den Soziologen Dorothy Swaine Thomas und William Isaac Thomas) zum Tragen, das schlicht besagt, daß Menschen ihre Handlungen nicht nach *der* Wirklichkeit ausrichten, wie sie *ist*, sondern nach *einer* „Wirklichkeit“, die so ist, wie diese Menschen jeweils glauben, daß sie ist. Damit kommen zwei neue bedeutende Aspekte ins Spiel: derjenige der Informiertheit und deren Umfang und Verlässlichkeit einerseits – nebst allem, was sich an dieser Stelle über Bildung, Aufklärung, Medien, *fake news*, Meinungsblasen, Quellenzugänglichkeit, etc. sagen ließe –, andererseits derjenige, den der Kommunikationswissenschaftler Paul Watzlawick „Interpunktion der Wirklichkeit“ genannt hat: Unsere Überzeugungen und Interessen sind mitbestimmend für alle unsere Wahrnehmungen der Wirklichkeit, also dafür, *wie* wir wahrnehmen und das dann interpretieren, und dafür, *was* wir überhaupt wahrnehmen (können).

Machen wir – als wäre die Angelegenheit noch nicht vertrackt genug – mit dem Biochemiker und Systemforscher Frederic Vester noch einen Schritt

²⁹ Georg Simmel: „Die Krisis der Kultur“, in: Ders.: *Der Krieg und die geistigen Entscheidungen. Grundfragen der Soziologie. Vom Wesen des historischen Verstehens. Der Konflikt der modernen Kultur. Lebensanschauung*, Frankfurt a.M. 1999, S. 46

³⁰ Mit genau dieser Begründung: „Denn der Reichtum ist nur als Mittel zu anderen Zwecken zu gebrauchen.“

(Aristoteles: *Nikomachische Ethik*, 1096a).

³¹ In den 1970er Jahren hat Elias mit großem Interesse die Ergebnisse der Forschungen in der Spieltheorie und in der Mustererkennung verfolgt und in seine Soziologie integriert.

³² Norbert Elias: *Über den Prozeß der Zivilisation*, Frankfurt a.M. 1997, Bd. 2, S. 324f

weiter in die Komplexität der Zusammenhänge:³³ Auch *Newtonsche Ursachen und Wirkungen* stecken gewissermaßen in einem Gefüge aus *Eliasschen Figurationen und Prozessen*. Bereits einzelne Verbindungen dieses Gefüges sind in der Regel nicht (und vielleicht niemals) *linear*, bei den meisten Zusammenhängen dürfte es nicht einmal genügend verlässliche mathematische Näherungsfunktionen geben. Innerhalb des Gesamtgefüges ergeben sich Rückkopplungen und komplexe und verschachtelte Regelkreise, was neben vielen anderen – manchmal schönen, manchmal fatalen – Effekten zu systematischen Kippunkten, Kollapsen und Schmetterlingseffekten führen kann, die, so vertraut uns dieses Vokabular aus Klima- und Corona-Diskussionen sein mag, sehr schwierig zu erkennen und noch schwieriger vorherzusagen sind. Bei der Orientierung und beim Agieren in dieser komplexen Wirklichkeit können uns auf Digitaltechnik basierende Maschinen in vielfältiger Weise unterstützen und uns Verrichtungen abnehmen, die uns überfordern oder die wir ungern tun oder die für uns gefährlich sind oder die eine Maschine schneller oder exakter oder weniger fehleranfällig auszuführen vermag, als ein Mensch dies könnte. Das führt dazu, daß diese Maschinen einen immer größeren und gestaltenden Raum in unserer Gegenwart einnehmen, in einem – allerdings ausschließlich übertragenen – Sinn wird davon gesprochen, daß sie *Macht* über uns erlangen. Das erzeugt sowohl Angst – weil das Verständnis dieser Maschinen viele Menschen überfordert und ihr Gebrauch ihnen gefährlich zu sein scheint, und in einigen Fällen ist er das auch zweifellos – als auch Hoffnung – weil Unangenehmes aus dem Lebensalltag von Menschen verschwinden, bedrohliche Fehler vermieden und insbesondere in Wissenschaften exaktere Ergebnisse gewonnen und schnelle Fortschritte erzielt werden können. Die mögliche *Ambivalenz der Auswirkungen der Anwendung* solcher Technologien wird hier deutlich. Auch wenn wir, wie ich im Folgenden begründen will, keinerlei rationalen Anlaß haben, zu be-

³³ Darauf können, nein, *dürfen* wir nicht verzichten: „Wir leben in einer Welt, deren ineinander greifende Abläufe für unseren menschlichen Geist schon immer schwer zu begreifen waren.“ Neben anderen Einflüssen haben technologische Entwicklungen „Wechselwirkungen so verdichtet, dass sie zu verstehen trotz aller wissenschaftlichen Erkenntnisse mit jedem Tag schwieriger zu werden scheint“, jedoch mit dem Effekt, „dass die sie verbindenden unsichtbaren Fäden hinter den Dingen für das Geschehen in der Welt oft wichtiger sind als die Dinge selbst.“: „Die Eigendynamik des Systems hat das Geschehen in die Hand genommen.“ Um die Kontrolle

fürchten, daß auf Digitaltechnik basierende Maschinen uns an Intelligenz zu übertreffen erlernen, die Menschheit unterdrücken und die Weltherrschaft übernehmen, besteht dringender Bedarf gesetzlicher Regulierungen für den Einsatz dieser Technologien, und die aktuellen Ansätze beispielsweise der EU sind auf jeden Fall zu begrüßen.

Um ein Verständnis dafür zu gewinnen, was wir von diesen Technologien zu befürchten haben und was wir hoffen dürfen, müssen wir klären, wovon wir hier „technisch“ eigentlich sprechen. Zwei mir besonders gravierend erscheinende Punkte möchte ich herausgreifen:

Erstens: Digitale Technik nimmt logische Verschaltungen von (in der Regel binären) Signalen vor. Damit lassen sich Rechnungen mit digitalen Zahlen abbilden, weshalb davon gesprochen werden kann, daß diese Maschinen *rechnen*, obschon es sich bereits hierbei genaugenommen um eine übersteigerte Metaphorik handelt. Mit *Lernen* oder gar mit *Intelligenz* hat das jedoch überhaupt und grundsätzlich nichts zu tun, denn nichts von dem, was eine Maschine in ihren logischen Schaltkreisen verarbeitet, hat *für die Maschine* irgendeine *Bedeutung*! Maschinen *lernen* nicht und *denken* nicht, sie *speichern* und sie *schalten* und gegebenenfalls *steuern* sie auch, allerdings weit entfernt davon, zu *entscheiden* und zu *handeln*.

Sowohl beim sogenannten „Maschinellen Lernen“ als auch bei der sogenannten „Künstlichen Intelligenz“ handelt es sich also zunächst einmal um ein terminologisches Mißverständnis bei der Begriffsbildung. Ein Meilenstein auf dem Denkweg hin zu diesem – meiner Ansicht nach höchst gefährlichen! – Mißverständnis ist Alan Turings These, Intelligenz mit ausreichend großem Täuschungspotential (für Lebewesen) *simulieren zu können*, sei dasselbe, wie *intelligent zu sein*. Gerade die aktuellen atemberaubenden Entwicklungen in der Digitaltechnik führen uns ja vor Augen, wie unsinnig diese These ist, denn was uns da – sei es vor Begeisterung oder vor Entsetzen – den Atem verschlägt, das sind die umfangreichen Datenmengen und die

über diese von *uns Menschen* entfesselte Dynamik nicht gänzlich zu verlieren, bedarf es „eine[r] neue[n] Sicht der Wirklichkeit“, müssen *wir Menschen* eine neue „Ebene unseres Denkens“ einnehmen und „uns auf ein neues Paradigma“ einstellen, „das sich an den auf unserem Planeten herrschenden Systemgesetzmäßigkeiten orientiert.“: „Um zu erfassen, was unsere Eingriffe in einem komplexen System bewirken, kommen wir nicht umhin, das Muster seiner vernetzten Dynamik verstehen zu lernen.“ (Frederic Vester: *Die Kunst vernetzt zu denken. Ideen und Werkzeuge für einen neuen Umgang mit Komplexität*, München 1999, S. 9f.

Geschwindigkeit und Exaktheit, mit denen sie verarbeitet und für den durchschnittlichen Menschen höchst schwierige Berechnungen ausgeführt werden können – aber nicht ein gewonnenes *Verständnis* und darauf aufbauende intelligente und kreative und vor allem selbstbewußte und eigenwillige *Ideen*, auch wenn manche Produkte dieser Technologien diesen (ersten) Eindruck erwecken mögen. Der Tübinger Informatik-Professor Torsten Grust spricht daher bezüglich der Diskussion um „Künstliche Intelligenz“ und „Maschinelles Lernen“ durchaus treffend von einer „Glorifizierung der Matrizenmultiplikation“ (und fügt (treffend) hinzu: „nur daß die Matrizen eben sehr groß sind“).

Wer daher meint, einen Digital-Computer mit einem Gehirn (eines Menschen oder eines anderen Tieres, das ist gleichviel) identifizieren zu können, der hat die Funktionsweise von mindestens einer der beiden Seiten seiner Gleichung nicht einmal ansatzweise verstanden. Möglicherweise werden Analog-, Bio- oder Quantencomputer eines Tages ein Gehirn tatsächlich simulieren können, wer weiß das schon (heute)? Daß und warum so etwas jedoch *mit digitaler* Technologie zu erreichen unmöglich ist und bleiben muß, das einzusehen, dazu bedarf es lediglich einer mengentheoretischen Betrachtung auf Mittelstufen-Gymnasialniveau. Denn:

Zweitens: Digitale Zahlen sind diskrete rationale Zahlen mit einer endlichen Zifferanzahl. Diese kann zwar theoretisch beliebig groß werden, bleibt dabei aber immer endlich. Bei der Erfassung der analogen Wirklichkeit durch digitale Technik – der also zurecht so genannten *Digitalisierung* – liegt daher mathematisch notwendig und folglich immer eine (doppelte) Reduktion der Wirklichkeitsdaten mit dem Faktor $\frac{1}{\text{abzählbar} \infty * \text{überabzählbar} \infty}$ vor.³⁴

Die Daten, die Digitaltechnologien verarbeiten, sind also an sich „klein“: Auch nur ein einziges Datum der realen Welt als Zahl *vollkommen exakt* erfassen zu wollen würde einen Speicherbedarf erfordern, den die Gesamtmenge der Elementarteilchen im Universum nicht hergibt, denn diese ist immer noch endlich.³⁵ Die digitale Verarbeitung von *Big Data* (große Datenmengen, nicht „große“ Daten) ist nur deshalb möglich, weil diese Data im Vergleich zur Wirklichkeit so armselig sind.

Das klingt despektierlich, ist es aber nicht. Um sy-

stemimmanente Wirkzusammenhänge zu erkennen, die, wie ich darzulegen versucht habe, bedeutsam sind, wenn wir uns Digitaltechnik für Mensch und Menschheit möglichst fruchtbringend nutzbar machen wollen, braucht es in der Regel nicht besonders exakte, sondern vor allem *viele* Daten.

Exemplarisch: Mustererkennung – etwa beim Versuch, mittels öffentlicher Kameras Straftäter oder Entführungsoffer aufzuspüren oder *Social Scoring*-Profile zu erstellen – funktioniert mit ungenauen – „fuzzy“ – Daten viel besser als mit genauen, so, wie wir beispielsweise in Londons National Gallery auf William Turners *Calais Pier* das wesentliche nicht erkennen, wenn wir zu nah davorstehen.

Die Reduktion der Wirklichkeit zum Behuf der Nutzbarmachung von Daten für Digitaltechnik ist also an sich unproblematisch, *solange* wir diese digitalisierte „Wirklichkeit“ nicht mit der Wirklichkeit verwechseln.

Hierzu ein weiterer wichtiger Aspekt: Über die mathematisch notwendige Datenreduktion hinaus ist zu bedenken, daß sich die große Mehrheit unserer Wahrnehmungen gar nicht quantifizieren läßt,³⁶ besonders, aber keineswegs ausschließlich, im Bereich der Gefühle, die wir empfinden. Und auf der anderen Seite der Erkenntnis denken und sprechen und erfassen wir die Welt keineswegs in deutlichen und klaren Begriffen, sie sind und bleiben verschwommen und tragen Unmengen an individuellen und nicht kommunizierbaren, nur sehr selten bewußten Konnotationen in einem Rucksack mit überabzählbar großem Fassungsvermögen mit sich herum auf der Reise durch unsere Sätze.

Ich erinnere mich, daß ich einmal als Kind, als ich krank war, auf die Frage meiner Mama, wie es sich anfühle, „so gelb“ geantwortet habe. Mir war auch in dem Moment bereits bewußt, daß das nicht verständlich sein könne, doch einen adäquateren Begriff hatte ich nicht und würde vermutlich auch heute keinen finden. Und meine Mama? Sie war, auch daran erinnere ich mich noch, mit meiner Antwort zufrieden und hat *etwas* verstanden.

Das führt den Gedankengang auf das kommunikative Substrat unseres menschlichen Welt- und Selbstverständnisses: Wenn wir miteinander sprechen, dann ist unser Einander-Verstehen immer nur

³⁴ Ich gestehe, daß diese Formulierung mathematisch nicht ganz sauber ist.

³⁵ Allein Gott, wie ihn etwa Gottfried Wilhelm Leibniz „versteht“, wäre wohl fähig, ein solches Meßergebnis zu erfassen.

³⁶ Kant widmet den ersten Abschnitt der *Disziplin der reinen Vernunft* in der *Kritik der reinen Vernunft* dem

Nachweis der Unmöglichkeit, die uns empirisch, d.i. mittels der Sinne gegebene Wirklichkeit mathematisch zu erfassen, denn: „Die Form der mathematischen Erkenntnis ist die Ursache, daß diese lediglich auf Quanta gehen kann.“ (*KrV*, A 714 (B 742)).

Qualität quantifizieren zu wollen, das ist und bleibt in jedem Kontext notwendigerweise ein Kategorienfehler.

Annäherung und hat immer ein grundlegendes tentatives Momentum. Ich glaube, daß wir anders – also mit exakten und klaren und deutlichen Begriffen (wie sie eine wirkliche „Künstliche Intelligenz“, jenseits der Metapher, womöglich benötigen würde) – gar nicht miteinander sprechen könnten, vermutlich hätten wir auch keinerlei Interesse daran. Unser Sprechen entspricht unserem Leben, in dem wir uns den großen Zielen, die wir verfolgen, auch immer nur annähern – zu unserem großen Glück –, und ohne jenes tentative Momentum würden wir uns auf dieses Abenteuer vermutlich gar nicht einlassen.

Diese Verschwommenheit bestimmt also unser menschliches Sprechen ebenso wie unsere anderen (nicht-sprachlichen) Handlungen und vermutlich alles, was wir über die Welt und über uns selbst und unser Leben denken. Was ich hier aber „*Verschwommenheit*“ nenne, das ist ganz etwas anderes als *Wahrscheinlichkeit* oder mathematische *Unge nauigkeit*, weshalb sie sich auch nicht stochastisch simulieren läßt.

Damit komme ich zurück zum radikalen Wandel des (erst esoterisch-wissenschaftlichen, dann, viel später, allgemein als selbstverständlich angenommen) *Weltbildes* in der frühen Neuzeit. Die philosophiehistorische Betrachtung dieser *Revolution* kann überaus erhellend sein, wenn wir – Laien wie Fachleute – versuchen wollen, einen adäquaten Blick auf unsere Gegenwart und in ihr auf die Herausforderungen und Chancen durch die neuen (zum Beispiel digitalen) Technologien zu gewinnen.

Ich hoffe, dieser *Gedankenrundgang* hat mindestens eines vor Augen geführt:

Je weiter wir uns mit unserem Selbstbild *als Menschen* in eine möglichst linear mathematisierte – hoch lebe der Dreisatz! – und in ihren Verflechtungen mechanisierte Weltfremdheit zurückkriechen, um so ebenbürtiger und schließlich überlegener müssen uns in dieser notwendig dystopischen Unwirklichkeit rechnende Maschinen erscheinen – und dann müssen sie uns ängstigen und erst recht ihre so wenig in den Diskussionen befragten menschlichen Schöpfer.

Versuchen wir so etwas wie ein Fazit:

Digitaltechnik, inclusive dessen, was sich heute so großspurig wie provokativ „Künstliche Intelligenz“ nennt, kann uns in unserem Miteinander-Leben in vielfältiger Weise und umfangreich unterstützen und uns vieles abnehmen, das wir als lästig empfinden oder schlicht der Menge oder der gefor-

derten Präzision wegen nicht oder nicht so gut bewältigen können. Diese Unterstützung jedoch ist grundsätzlich nur dort möglich, wo wir uns relativ gefahrlos mit dem mathematisch-mechanistischen Weltbild und den auf digitale Zahlen reduzierten Wirklichkeitsdata *zufriedengeben* können.

Wir tun aber gut daran, uns bewußt zu machen und zu bleiben, daß die „in der Sprache der Mathematik“ *beschriebene Welt*, so *einfach und schön* sie auch immer uns erscheinen mag, mit der wirklichen Wirklichkeit und einem echten gelebten Leben nicht einmal annäherungsweise und tentativ auch nur vergleichbar ist und niemals sein wird.³⁷ Das mögen nun die eine oder der andere enttäuschend finden – ich wüßte freilich nicht recht, *warum?*

(Göttingen, 16. Juni 2023)

* * *

Warum die Kombination aus künstlicher und natürlicher Dummheit viel gefährlicher ist als jede Art von Künstlicher Intelligenz

Paul Lukowicz

Ich kann kein Wort Japanisch, habe auch keine Ahnung, was die einzelnen Zeichen bedeuten. Stellen Sie sich einmal vor, ich würde mir die Mühe machen, sämtliche japanische Zeitungsberichte, Schulbücher usw. der letzten 20 Jahre durchzusehen und mir daraus eine Statistik zu erstellen, die es mir erlaubt (ohne an meiner fehlenden Kenntnis von Japanisch etwas zu ändern), für eine jede beliebige Folge von japanischen Zeichen vorherzusagen, welches Zeichen (oder welche Zeichenfolge) höchstwahrscheinlich danach folgt. Stellen Sie sich nun vor, jemand würde mir auf Japanisch eine Frage stellen. Die Frage wäre für mich einfach eine Folge nichtssagender Zeichen. Aber ich könnte auf meine Statistik zurückgreifen und das nächste am wahrscheinlichsten zu erwartende Zeichen (oder die nächste Zeichenfolge) vorhersagen. Damit hätte ich eine neue, ein wenig längere Folge von Zeichen. Ich könnte darauf wieder meine Statistik anwenden, um eine noch längere Folge zu bekommen usw., bis ich einen langen „Antworttext“ produziert hätte. Würden Sie, wenn sich der so entstandene Text plausibel und sprachlich korrekt anhören würde, sagen, dass ich (1) Japanisch gelernt habe, (2) die Frage verstanden habe oder gar (3) ein Verständnis von der Domäne habe, auf die sich die Frage bezogen hat? Wohl kaum!

auf der Illusion, die Wirklichkeit bestehe aus Matrizen und ihre Zukunft aus Multiplikationen derselben. Wischen wir der Menschheit Sand aus unseren Augen!

³⁷ Im Anschluß an Grust könnte gesagt werden: Alle mit „Künstlicher Intelligenz“ verbundenen Ängste und die meisten der mit ihr verbundenen Hoffnungen beruhen

Interessanterweise wird derzeit darüber spekuliert, ob die Tatsache, dass sogenannte generative KI-Systeme, die auf der Large Language Models (LLMs)-Technologie basieren (z.B. ChatGPT), die exakt nach demselben Prinzip funktionieren, nicht nur als wirklich intelligent bezeichnet werden können, sondern sogar Bewusstsein entwickelt haben könnten. Insbesondere ist die alte Diskussion wieder entbrannt, ob und inwieweit sich KI-Systeme „verselbstständigen“ können, inklusive alarmistischer Warnungen über eine mögliche Auslöschung bzw. Versklavung der Menschheit durch außer Kontrolle geratene KI, die sich selbst zur „dominierenden Spezies“ aufschwingen will. Gleichzeitig steigen die Erwartungen an das, was solche Systeme leisten sollten, ins Unermessliche. Die Nutzerbasis von ChatGPT ist innerhalb kürzester Zeit auf Hunderte von Millionen angestiegen. Die Anwendungen reichen von der Lösung von Hausaufgaben in so ziemlich allen Fächern, über die Verfassung von Abschnitten wissenschaftlicher Arbeiten, das Erstellen von Computerprogrammen bis hin zur Liebesberatung. Die Integration von Bildrepräsentation erlaubt darüber hinaus die Interpretation komplexer Szenen aus der realen Welt so wie die Erzeugung realistischer Bilder oder gar Videos aus einfachen textuellen Beschreibungen.

Was sollten wir von der Entwicklung halten? Nun, was die Furcht von der bössartigen „Superintelligenz“ anbetrifft, so ist hier ein Zitat des berühmten Science Fiction-Autors Arthur C. Clarke angebracht: „Any technology that is advanced enough is indistinguishable from magic“. Ist aber eben keine Magie, wenn man sie erst einmal versteht. Ähnlich könnten wir sagen: „Any statistical model that is advanced/complex enough is indistinguishable from intelligence“. Ist aber eben nur Statistik bzw. eine Folge von Vektor-Matrix-Multiplikationen und einfacher nichtlinearer Funktionen. Werden wir eines Tages genug über das menschliche Gehirn, das Wesen von Intelligenz und das Bewusstsein wissen, um Maschinen zu bauen, die uns in allen Aspekten der Intelligenz ebenbürtig oder gar überlegen sind; vielleicht sogar solche, die ein eigenes Bewusstsein und einen freien Willen entwickeln können? Ich weiß es nicht. Sollten wir je soweit kommen, dann müssen wir auf jedem Fall dieselben strengen ethischen Grundsätze und Beschränkungen anwenden, wie wir es heute aus guten Gründen z.B. in Bezug auf das Klonen von Menschen tun. Dass aus der heutigen Entwicklung in LLMs und generativer AI eine potentiell gefährliche „Superintelligenz“ entsteht ist allerdings in etwa genauso wahrscheinlich, wie dass morgen

eine Flotte außerirdischer Raumschiffe im Sonnensystem erscheint, um die Erde anzugreifen.

Die eigentliche Gefahr ist nicht, dass die vorhandenen Systeme zu viel Intelligenz entwickeln, sondern vielmehr, dass Ihre Möglichkeiten maßlos überschätzt werden. Das Problem ist, dass Menschen ihnen zu sehr (teils blind) vertrauen und sie unbedacht in Bereichen und in einer Art und Weise angewendet werden, die potentiell gefährlich und schädlich ist. Angesichts der erstaunlichen Ergebnisse, die Systeme wie z.B. ChatGPT immer wieder produzieren, liegt eine solcher Überschätzung nahe. Man sollte aber niemals vergessen, dass die Systeme kein wirkliches Verständnis haben, sondern einfach nur die Statistik der Texte, auf denen sie trainiert wurden, replizieren. Dies bedeutet nicht nur, dass die Ergebnisse oft genauso falsch sind wie sie sprachlich autoritativ und geschmeidig formuliert sind. Die statistische Natur der Modelle führt auch dazu, dass all die Unausgewogenheit und Vorurteile, von denen die digitale Welt geprägt ist, weitgehend ungefiltert übernommen werden. Solche Modelle sollten in keinem Fall ohne geeignete Vorkehrungen als Grundlage für Anwendungen dienen, die menschliche Schicksale beeinflussen können (z.B. Bewertung der Kreditwürdigkeit, Personalauswahl, öffentliche Sicherheit, Justiz usw.). Die Frage, wie solche Vorkehrungen aussehen können, ist ein aktuelles Forschungsthema im Bereich Fair/Trustworthy/Explainable AI und bisher bei weitem nicht gelöst. Auch die Gefahr für demokratische Strukturen, die durch KI generierte Desinformation und eine weitere Verstärkung des „echo chamber“-Effektes entsteht, darf auf keinem Fall unterschätzt werden. Zuletzt sollte man auch nicht übersehen, dass es gar nicht notwendig ist, dass sich Systeme zu einer „Superintelligenz“ entwickeln und selbständig machen, damit ihre Aktionen unvorhersehbar werden. Ganz im Gegenteil. Es ist allgemein bekannt, dass insbesondere moderne Deep Learning Verfahren unter dem „black box“-Problem leiden. Aus riesigen Datensätzen, die von Menschen nicht mal ansatzweise zu überblicken sind, werden durch komplexe, teils probabilistische Optimierungsverfahren aus hunderten Milliarden von Matrixelementen bestehende Repräsentationen generiert. Daraus abzuleiten, was ein solches System in welcher Situation „machen“ wird, ist kaum möglich. Hinzu kommt, dass solche Systeme nicht einzeln für sich isoliert agieren, sondern dynamisch miteinander und mit Menschen kommunizieren und interagieren. Bei der Komplexität und Dynamik, die hier im Spiel ist, muss man mit unvorhersehbaren emergenten Effekten rechnen, bis hin

zum chaotischen Verhalten. Gleichzeitig übertragen wir solchen Systemen die Kontrolle über mehr und mehr Bereiche unseres Lebens, inklusive kritischer Infrastrukturen wie Verkehr, Energie, Kommunikation, Finanzmärkte usw. Somit könnte es in der Tat sein, dass die KI eines Tages für katastrophale, gar existenzbedrohende Ereignisse „verantwortlich“ sein wird. Aber nicht, weil sie zu intelligent ist, sondern weil sie nicht intelligent genug ist und weil sie in einer Art und Weise verwendet wird, die nicht ausreichend durchdacht ist. Unser Untergang wäre dann nicht die Künstliche Intelligenz, sondern eine Kombination aus künstlicher und natürlicher Dummheit. –

Abschließend ist es wichtig, darauf hinzuweisen, dass trotz all der erläuterten Probleme die Technologie durchaus als eine Revolution in der KI angesehen werden kann. Sie hat enormes Potential, die Leistungsfähigkeit von KI-Systemen auf ein neues Level zu heben und eine Fülle von neuen Anwendungen zu eröffnen. Zum ersten Mal in der Geschichte der KI können wir nun Systeme bauen, die ein umfassendes Modell von der Welt und der Art und Weise, wie Menschen die Welt wahrnehmen und bewerten, haben. Damit könnten fundamentale Probleme der KI wie das berühmte „Grounding Problem“ gelöst und Systeme gebaut werden, die nicht nur blind Fragen beantworten, sondern auch konstruktiv mit Menschen zusammenarbeiten können. Dafür müssen wir aber sicherstellen, dass die Arten und Weisen, wie KI entwickelt und vor allem genutzt wird, genau durchdacht und an bestimmten Stellen auch gesetzlich reguliert werden. Hierfür sind geeignete interdisziplinäre Ansätze nötig, die auch in der Ausbildung des KI-Nachwuchses berücksichtigt werden müssen. Und wir dürfen niemals vergessen, dass diese Modelle Approximationen sind, die nur so gut sind wie die Daten, aus denen sie entstanden sind.

(Kaiserslautern, 19. Juni 2023)

KORREKTUR

In der vorhergehenden Ausgabe gab es einen Tippfehler: Der Artikel von Marco Baschera trug versehentlich die Jahreszahl 2013 anstelle der – korrekten – aktuellen, so daß der verwirrende Eindruck entstand, der Text sei bereits zehn Jahre alt und für den *Rundbrief* (für den er tatsächlich geschrieben wurde) nicht aktualisiert worden. Verantwortlich

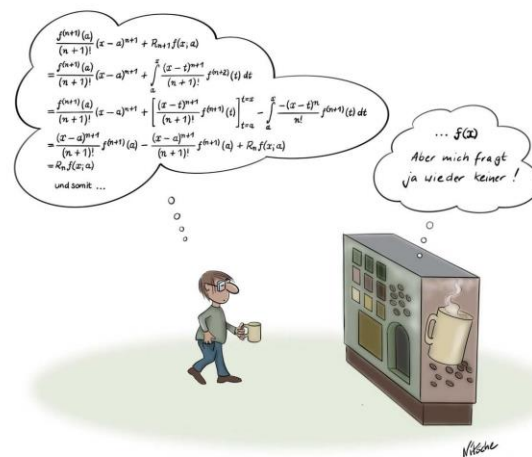
* Bitte kennzeichnen Sie an redaktion@aphin.de gerichtete Einreichungen für diese Rubrik als Leserinnen- oder Leserbriefe. Die Redaktion behält sich die Entscheidung

für den Fehler ist allein der Herausgeber, der hiermit um Entschuldigung bittet.

LESERINNEN- UND LESERBRIEFE*

...stellen eine Möglichkeit dar, die Rundbriefe auch zu einem Medium des *Gedankenaustausches* zu machen und zugleich Autorinnen und Autoren eine Rückmeldung über ihre jeweils dargelegten Positionen zu geben. Das gilt für diese erste Ausgabe mit einem Schwerpunkt in ganz besonderem Maße, denn die Ausführlichkeit und die Vielfalt, mit der hier das Thema betrachtet wurde, birgt wohl ein Potential, als Boden für weitere Reflexionen und Diskussionen zu dienen. Ihre Gedanken, liebe Leserinnen und Leser, zu den auf diesen Seiten formulierten Gedanken der Autorinnen und Autoren zu erfahren, ist daher seitens der Redaktion ausdrücklich erwünscht.

ANSICHTSSACHE



ZITATE

„Wer sich durch die autoritären Stimmen der Vergangenheit in seinem Vermögen, *für sich selbst zu denken*, beirren läßt, ist verloren.“

(Norbert Elias)

„Wie man die Welt einst wahrzunehmen gelernt hat, sie gesehen, gehört, geschmeckt, gerochen hat, mit wie viel Freude oder vielleicht auch Schmerz, so sieht und riecht und hört man im Leben weiter.“

(Tanja Maljartschuk)

über die Veröffentlichung vor. Eventuelle Änderungen am Text erfolgen hingegen nicht ohne Absprache mit Autorinnen und Autoren.

RÄTSEL**

Ein Rätsel

Er donnerte auf die Erde, vielleicht.

Als Spore, vielleicht.

Er ist kein Tier und gar nicht grün.

Sein Bett nicht aus Metall und manchmal nur von Holz.

Der Freunde hat er gar sehr viele,
gleich der Fische im Meer sich ganz verschieden.
Kein Museum je gesehen und doch so oft schon dort gewesen.

(Maximilian Nüßler, Göttingen)

Auflösung des vorhergehenden Rätsels

Die Auflösung des vorherigen Rätsels, wer dasjenige dieser Ausgabe stellen würde, ist bereits gegeben. Interessanterweise entstand dieses Rätsel nicht mit Blick auf den *Rundbrief*, sondern im Rahmen eines Seminars zum Thema Kreativität. Zudem steht es in einem ganz losen Zusammenhang zu einer Erzählung von Javier Marías („Unfertige Gestalten“), über den in der vergangenen Ausgabe einiges zu lesen war.

Die *Einladung und Bitte* an Leserinnen und Leser, Rätselvorschläge (nebst Lösungen) für kommende Ausgaben des *Rundbriefs* einzureichen, bleibt bestehen. Der Herausgeber würde diese, den Rückmeldungen nach beliebige Rubrik gerne erhalten, sieht sich jedoch in der nächsten Zeit nicht in der Lage, sie selbst zu gestalten.

VERANSTALTUNGEN

APHIN V 2023 – Natur, Kultur und Technik

Vom 23. bis 25. Juni 2023 findet an der Mosel unsere fünfte, öffentliche **APHIN-Tagung** statt. 27 Vortragende aus unterschiedlichen Fachbereichen werden über das Tagungsthema **Natur, Kultur und Technik** referieren. Wir laden Sie herzlich ein, teilzunehmen und mitzudiskutieren. Tagungsprogramm und Anmeldeformular finden sie auf unserer Homepage unter dem Menüpunkt Veranstaltungen/Tagungen.

Am Abend des 23. Juni 2023 feiern wir zudem gemeinsam das zehnjährige Bestehen von APHIN. Wir freuen uns darauf, Sie auf diesen Veranstaltungen wiederzusehen.

(jhf)

Leseseminar in Esthal (Pfalz):

Carl Friedrich von Weizsäcker: *Der Mensch in seiner Geschichte*

Carl Friedrich von Weizsäckers 1991 erschienenenes

Buch *Der Mensch in seiner Geschichte* – ein konziser „Rundritt“ durch das Denken des Physikers und Philosophen – wird Gegenstand des achten Leseseminar im Kloster St. Maria in Esthal sein, das vom Freitag, 21. Juli (14.00 Uhr) bis Sonntag, 23. Juli (13.00 Uhr) stattfindet. Vorwissen wird nicht erwartet, lediglich die Bereitschaft, sich konzentriert und undogmatisch auf die Lektüre einzulassen. Eine detailliertere Beschreibung der Seminarinhalte finden Sie in der vorherigen *Rundbrief*-Ausgabe (Nr. 29, Mai 2023).

Für das Seminar wurden im Kloster bereits Zimmer vorab gebucht. Die Anmeldung zum Seminar kann formlos an die Adresse redaktion@aphin.de erfolgen. Bitte geben sie an, ob Sie ein Standard- oder ein Komfortzimmer wünschen. Rustikale Verpflegung ist im Übernachtungsangebot inbegriffen. Weitere Informationen zum Kloster finden Sie unter <https://kloster-esthal.de/>.

Zwecks Planungssicherheit und wegen der beschränkten Platzanzahl wird um eine möglichst zeitnahe Anmeldung gebeten.

Alle Teilnehmer und Teilnehmerinnen erhalten ein Teilnahmezertifikat. Bei Fragen wenden Sie sich bitte gerne an redaktion@aphin.de.

4. APHIN-Symposium: *Kant im Kontext der europäischen Aufklärung*

Das 4. APHIN-Symposium wird sich dem Königsberger Philosophen Immanuel Kant widmen und unter dem Titel *Kant im Kontext der europäischen Aufklärung* sein Leben und sein Werk im Licht seiner und unserer Zeit zu betrachten unternehmen. Dabei soll Kant nicht ausschließlich als *Denker der Aufklärung*, sondern gleichermaßen auch als *Denker in der Aufklärung* angesehen werden.

Das Symposium wird, abermals beginnend mit dem *Welttag der Philosophie*, vom Freitagabend des 16. bis zum Sonntagnachmittag des 19. November 2023 wie gehabt in Göttingen und in Kooperation mit dem Philosophischen Seminar der Georg-August-Universität stattfinden.

Eine detailliertere Beschreibung der angestrebten Inhalte des Symposiums finden Sie in der vorherigen *Rundbrief*-Ausgabe (Nr. 29, Mai 2023).

ARBEITSGRUPPEN

Philosophie und Informatik

Kürzlich ist unsere Arbeitsgruppe nach einer längeren Pause endlich wieder aktiv geworden. Die Arbeitsgruppe besteht seit 2018 und hat von Anfang an großes Interesse erfahren. Ihre Mitglieder

redaktion@aphin.de senden, vorzugsweise mit einem Hinweis, wie Sie auf die Lösung gekommen sind.

** Es gibt bei diesen Rundbrief-Rätseln zwar nichts zu gewinnen, dennoch können Sie gerne Ihre Lösung an

kommen aus verschiedenen Fachbereichen und bringen vielfältige Kompetenzen mit. Gemeinsam haben wir viele Themen im Bereich Künstliche Intelligenz bearbeitet und diskutiert. Aus verschiedenen Gründen hat sich jedoch die Intensität unserer Aktivitäten allmählich verringert, bis sie schließlich zum Stillstand kam. Eine Wiederbelebung stand immer im Raum, aber es fehlte das passende Thema. Glücklicherweise kam uns der Zufall zu Hilfe durch Dr. Reto Schöllys vor einigen Monaten veröffentlichtes Buch *How I Became A ChatBot*, das sich als geeignetes Thema für unsere Arbeitsgruppe erwies.

Ein Anruf mit Herrn Schölly brachte dann schnell alles notwendige auf den Weg und die Mail an die Teilnehmer der Arbeitsgruppe konnte verschickt werden. 15 Personen haben sich direkt zurückgemeldet und ihr Interesse an einer Bearbeitung des Themas ChatGPT bekundet. Am Montag, dem 5. Juni, fand die erste Online-Sitzung einer geplanten Reihe zum Thema statt. In dieser ersten Veranstaltung stellte Herr Schölly sich und sein Buch vor. Diese Reihe wird nun an jedem ersten Montag eines Monats fortgesetzt. Den Teilnehmern wird im Voraus ein Kapitel des Buches als Diskussionsgrundlage zugeschickt, das dann in der eigentlichen Veranstaltung besprochen wird. Vorläufig sind fünf Termine geplant.

Bei Interesse an dem Thema und der Arbeitsgruppe setzen Sie sich bitte über die folgende Email-Adresse mit Herrn Dipl. Inf. Ralf Kierspel in Verbindung: [philosophie-informatik\(at\)aphin.de](mailto:philosophie-informatik(at)aphin.de).
(rk)

LITERATURHINWEISE***

Anstelle von Literaturhinweisen finden Sie in dieser Ausgabe eine Einladung zur Mitwirkung. Seitens der Redaktion wird angestrebt, in dieser Rubrik zukünftig möglichst wenige – letztlich lediglich werbende – Verlagstexte zu reproduzieren, die Sie auf jeder Verkaufsplattform für Bücher im Internet finden können. Stattdessen sind Besprechungen von Büchern ausdrücklich erwünscht, die Leserinnen und Leser der Rundbriefe für andere Leserinnen und Leser der Rundbriefe verfassen. Dabei sind dann auch größere Gestaltungsspielräume „erlaubt“ und wird deren Ausnutzung sogar begrüßt, als dies bei trockenen Rezensionen wissenschaftlich relevanter Literatur allgemein üblich ist – solange der Pfad der Sachlichkeit nicht vollständig verlassen wird.

*** Eine vollständige Liste aller bislang hier erschienenen Literaturhinweise finden Sie auf der Rundbrief-

Die Redaktion würde sich ausgesprochen freuen, wenn sich im Laufe der Zeit (und der kommenden Ausgaben) ein kleiner Kreis von Personen zusammenfände, von denen jede bereit ist, vielleicht einmal im Jahr eine Besprechung für die Literaturhinweise des Rundbriefs zu verfassen. Titel können eigenen Interessen folgend gewählt sein, sofern sie dem weiten Themenfeld des APHIN nicht gänzlich fern liegen. Die Redaktion macht hier aber auch gerne einige Vorschläge.

Sollten Sie an der Mitwirkung an diesem Projekt interessiert sein, so melden Sie sich bitte bei der Redaktion über die untenstehende E-Mail-Adresse. Vielen Dank!

IMPRESSUM

Verantwortlich für die Inhalte dieses Rundbriefs ist, sofern Beiträge nicht anders namentlich gekennzeichnet sind, Torsten Nieland. Hinweise auf eventuell fehlerhafte Angaben werden jederzeit gerne entgegengenommen.

APHIN e.V.
An der Krone 1
56850 Enkirch / Mosel

redaktion@aphin.de
www.aphin.de

ISSN: 2748-3711

Die Redaktion möchte Leserinnen und Leser des Rundbriefs ausdrücklich zur Mitwirkung an kommenden Ausgaben ermuntern. Hinweise hierzu entnehmen Sie bitte dem Informationsblatt auf der Rundbrief-Seite unserer Homepage.

Seite unserer Homepage. Bitte verstehen Sie diese Liste auch als Anregung zu eigenen Beiträgen.