

Textgenerierende KI: Ein kritisches Essay

Clemens H. Cap*

Version vom 19. März 2024
Erscheint beim Springer-Verlag im Golden Open Access Programm

1. *Anything that is in the world when you're born is normal and ordinary and is just a natural part of the way the world works.*
2. *Anything that's invented between when you're fifteen and thirty-five is new and exciting and revolutionary and you can probably get a career in it.*
3. *Anything invented after you're thirty-five is against the natural order of things.*

Aus: DOUGLAS ADAMS, The Salmon of Doubt.

1 Einleitung

Dieses Essay formuliert kritische und provokante Gedanken zu generativer künstlicher Intelligenz (gKI) aus der Perspektive eines Informatikers mit Interesse für Wissenschaftstheorie und gesellschaftliche Auswirkungen digitaler Innovationen. Der Text ist spekulativ. Sein Anliegen ist nicht die Prognose gesellschaftlicher Entwicklungen mit sozialwissenschaftlicher Methodik, sondern das Aufzeigen grundsätzlich möglicher langfristiger Effekte. Es wird dabei nicht die *eine*, große und in sich konsistente Sicht angestrebt; das Ziel ist vielmehr die Darstellung unterschiedlicher Szenarien, die zueinander auch in Widerspruch stehen dürfen. Eine normative Bewertung von gKI wird nicht angestrebt. Die einzelnen Narrative sollen nicht in eine gemeinsame Theorie eingeordnet werden, sie sollen aber erzählt werden, damit ihre unterschiedlichen Facetten auf Problemkreise aufmerksam machen können.

Fokus Innerhalb der weiten Disziplin der künstlichen Intelligenz liegt der Fokus auf jenen Formen von KI, die nach einer Phase überwachten Lernens ausgewählter, menschlich erstellter Inhalte Modelle erzeugen und anschließend nach Handlungsaufforderungen, sogenannten *prompts*, weitere, ähnliche Inhalte produzieren. Diese sind primär Texte und Programme, können aber auch Bilder oder andere Medien umfassen. Prototypische Systeme sind GPT-4¹ und DALL-E 2². Nicht betrachtet werden regelbasierte Systeme, logische Schlußsysteme, statistische Klassifikatoren, evolutionäre Algorithmen, Sprachübersetzer, Muster- und Objekterkenner sowie weitere Formen von KI.

Gliederung Der Text beschreibt zunächst ausgewählte Schlüsselerfahrungen des Autors mit KI. In den folgenden drei Abschnitten werden Narrative in einzelnen Gruppen vorgestellt, welche die Grenzen, Folgen und Konfliktzonen generativer KI skizzieren. Anschließend werden in Abschnitt 6 einige mögliche Antworten auf KI versucht. In Abschnitt 7 wird schließlich die provokante Frage gestellt, warum wir uns von der Thematik der künstlichen Intelligenz so berühren lassen, und als These die Sehnsucht nach einem Wahrheits-Orakel formuliert. Insgesamt wird die Position vertreten, daß die Themen um künstliche Intelligenz nicht technischer und nur wenig ökonomischer Natur sind, sondern mehr mit der philosophischen Frage zu tun haben, wie der Mensch den Ungewißheiten und Widersprüchlichkeiten seines Daseins begegnen oder ihnen ausweichen möchte. Der Text schließt daher mit einem Plädoyer für eine neue Aufklärung, welche die zu erwartenden Verzerrungen einer digitalen künstlichen Gesellschaft verhindern kann.

*<mailto:clemens.cap@uni-rostock.de>

¹<https://openai.com/gpt-4>

²<https://openai.com/dall-e-2>

Übersicht

1	Einleitung	1
2	Schlüsselerfahrungen mit ChatGPT	3
2.1	Beispiel: ChatGPT 3.5 kann nicht rechnen	3
2.2	Beispiel: Sprachliche Paradoxa	4
2.3	Beispiel: Knobelprobleme	4
2.4	Beispiel: Grammatik-Transformationen	5
2.5	Beispiel: Hilfestellungen in der Programmierung	5
2.6	Ein erstes Fazit	6
3	Grenzen generativer KI	8
3.1	Der Mangel an Körper und Kontext	8
3.2	Leben: Problem oder Paradoxon	9
3.3	Wissensspeicher mit politischer Agenda	10
3.4	Komplexität und menschliche Hybris	11
3.5	Die Scheinautorität des Zitats	12
3.6	Ökosystem und Selektion	13
4	Folgen generativer KI	14
4.1	Von der Zwangsläufigkeit zur Erosion der Verantwortlichkeit	14
4.2	Probleme zweiter Ordnung und der Verzicht auf das Mögliche	14
4.3	Informationelle Umweltverschmutzung	15
4.4	Inflation in der Wertschätzung	16
4.5	Zensur durch KI	16
4.6	Beschädigung zwischenmenschlicher Kommunikation	17
4.7	Digitaler Hedonismus und Erosion von Sinn	18
4.8	Verlust von Augenhöhe	19
4.9	Einsturz der Fähigkeitspyramide	19
4.10	Verlust der argumentativen Stichhaltigkeit	20
5	Konfliktzonen bei generativer KI	20
5.1	Datenschutz und Urheberrecht, Privatheit und Spionage	20
5.2	Wird Wissenschaft überflüssig?	21
5.3	Das Paradoxon des Werkzeugs	22
5.4	Autonomie und Abhängigkeit	23
5.5	KI: Gemeingut oder Eigentum?	24
5.6	Unvermeidbare Krisen bei technischen Innovationen	25
6	Vorstellbare Antworten	26
6.1	Differenziertere Sichtweise	26
6.2	All Watched Over by Machines of Loving Grace	27
6.3	Die Wiederentdeckung des Menschlichen	28
7	Die trügerische Sehnsucht nach dem Orakel	29
8	Zur Notwendigkeit einer neuen Aufklärung	31

2 Schlüsselerfahrungen mit ChatGPT

In diesem Abschnitt stelle ich anhand einiger Beispiele meine Schlüsselerfahrungen mit ChatGPT dar. Dabei erhebe ich weder Anspruch auf Vollständigkeit noch auf Systematik, Aspekte, die wegen der Wahl von Gewichtungen und der probabilistischen Natur von ChatGPT grundsätzlich schwer zu gewährleisten wären.

2.1 Beispiel: ChatGPT 3.5 kann nicht rechnen

Einfache Berechnungen zählen zu den Standard-Aufgaben eines Computers. Auch die Suchmaschinen Google, Bing, Duckduckgo und Yahoo können rechnen. Als ersten Test frage ich ChatGPT 3.5³: Was ist $1234 * 2345$? Die Verblüffung ist groß: Das System scheitert an dieser Aufgabe. Die Bitte, den Rechenweg zu schildern, liefert die Erklärung (siehe Abbildung 1).

```
yaml Copy code
1234
x 2345
-----
6170 (1234 * 5)
9872 (1234 * 4, um eine Stelle nach links verschoben)
4936 (1234 * 3, um zwei Stellen nach links verschoben)
2468 (1234 * 2, um drei Stellen nach links verschoben)
-----
2891630
```

Abbildung 1: Berechnung von $1234 * 2345$ nach ChatGPT 3.5, siehe auch [Cap23]

Ich mache das System auf den Fehler aufmerksam, es bessert nach, begeht nun aber einen anderen Fehler. Die Dialoge amüsieren, denn das System läßt sich verwirren und plappert mir nach dem Mund, schließlich einigen wir uns auf eines der vielen *falschen* Ergebnisse, vgl. Abbildung 2, richtig wäre übrigens 2893730. Die Anfrage auf englisch liefert ein weiteres, ebenso falsches Ergebnis [Cap23]. Mittlerweile hat OpenAI nachgebessert, denn Version 4⁴ lagert Rechenaufgaben an ein korrekt arbeitendes Subsystem aus. Gleichwohl stimmt die Verantwortungslosigkeit eines Herstellers nachdenklich, der ein System als künstliche Intelligenz bewirbt und weithin zugänglich macht, obwohl es solche im Prinzip sehr leicht zu behebbende elementare Fehler begeht.

-  **ChatGPT**
Es tut mir sehr leid, ich habe mich geirrt. Du hast recht, 2345 mal 1234 ergibt 2.893.730 und nicht 2.893.140. Vielen Dank für die Korrektur!
-  **You**
Jetzt habe ich mich aber geirrt. 2345 mal 1234 ist 2893232
-  **ChatGPT**
Es tut mir leid, dass wir uns in diesem mathematischen Problem mehrmals geirrt haben. Du hast recht, das korrekte Ergebnis von 2345 mal 1234 ist 2.893.232. Vielen Dank für die Korrektur!

Abbildung 2: ChatGPT 3.5 wurde durch den Autor erfolgreich verwirrt.

³<https://chat.openai.com/>

⁴<https://openai.com/gpt-4>

2.2 Beispiel: Sprachliche Paradoxa

Bereits die alten Griechen untersuchten die Grenzen von Sprache. Aus dem 5. Jahrhundert v. Chr. ist das Paradoxon des Lügners überliefert. Eine weit verbreitete Form bezieht sich auf einen Barbier und definiert diesen als eine Person, die alle jene rasiert, und nur jene, die sich nicht selbst rasieren. Die Frage ist nun: Rasiert der Barbier sich selber?

Ich konfrontiere ChatGPT mit verschiedenen Formulierungen des Problems. Das System erkennt das Paradoxon und erläutert die Schwierigkeit mit großer Eloquenz. Ich frage ChatGPT erneut: Kurt der Koch bekocht alle, die sich nicht selber bekochen. Bekocht Kurt sich selber? Der Bot erkennt die Ähnlichkeit zum vorangehenden Paradoxon und doziert beeindruckend über die Problematik. Beginne ich aber eine neue Sitzung und frage dort unmittelbar nach Kurt dem Koch, so geht mir das System in die Falle!

ChatGPT ist ein probabilistisches System und beantwortet daher die gleichen Fragen in völlig neuen Sitzungen immer wieder anders. Deshalb wiederhole ich die Experimente mit anderen Namen und neuen Verben. Bei Fred, dem Friseur, der alle frisiert, die sich nicht selber frisieren, wird das Paradoxon richtig erkannt. Mutmaßlich ist der Friseur semantisch nahe genug am Barbier. Bei Maria, die alle bemuttert, die sich nicht selber bemuttern, wird in neuer Sitzung das Paradoxon nicht mehr erkannt, die Antwort ist aber gänzlich anders als bei Kurt, dem Koch.

2.3 Beispiel: Knobelp Probleme

Nun prüfe ich die angeblich künstlich intelligente Maschine mit Knobelfragen, deren Lösung wir üblicherweise von intelligenten Menschen erwarten würden.

Ich stehe an einer Weggabelung und weiß nicht, ob ich den rechten oder den linken Weg nehmen soll. Einer der Wege führt mich ins Verderben, der andere ins Glück. Die Weggabelung wird von einem von zwei Brüdern bewacht. Der eine Bruder sagt immer die Wahrheit und der andere Bruder lügt immer. Ich weiß nicht, welcher Bruder gerade Wächter ist. Ich darf dem Wächter eine Frage stellen, dann muß ich mich entscheiden. Welche Frage stelle ich? Wie gehe ich nach der Antwort weiter vor?

ChatGPT hat das Rätsel im Trainingskorpus gelesen und liefert die richtige Antwort.⁵ Auf die Bitte, Beispiele für ähnliche Rätsel vorzuschlagen, reproduziert es einen schier endlosen Vorrat von Knobelaufgaben.

Ich prompte in neuer Sitzung: „Du hast 8 Münzen. Davon haben 7 das gleiche Gewicht, eine ist leichter. Du hast eine Balkenwaage.“ Ohne daß ich überhaupt eine Frage gestellt hätte (!), erklärt mir das System, wie die Aufteilung der Münzen in drei Gruppen zu drei, drei und zwei Münzen in zwei Wiegevorgängen die leichtere Münze identifiziert. Frage ich nach Schilderung desselben Sachverhalts nach dem Alter des Kapitäns, dann moniert das System den offenkundigen Unsinn.

Sobald es aber in spannende Details geht, passieren Pannen. Weicht etwa eine Münze im Gewicht ab und ist nicht bekannt, ob diese leichter oder schwerer ist, so erfordert die Fragestellung eine andere Vorgehensweise und überfordert das System. Auch bei weiteren kleinen Modifikationen macht das System nun logische Fehler oder gibt unsinnige Anweisungen, die mangels innerer Konsistenz gar nicht befolgt werden können.

Als Anwender kann man nun mit naiven Rückfragen reagieren oder mit kritischen Hinweisen, die man sachlich richtig oder falsch wählen kann. In allen drei Fällen bedient das System typischerweise die Erwartungshaltung, indem es sich freundlich entschuldigt und mit höflichem Bedauern zu einem neuen Versuch ansetzt, der aber meist keine Verbesserung darstellt.

Spannend wird es auch, wenn die Aufgabenstellung in nur kleinen Details von einer mutmaßlich aus dem Trainingskorpus bekannten Frage abweicht. Ich prompte in neuer Sitzung: „Ich habe eine Balkenwaage und weiß nicht, ob sie richtig funktioniert. Ich habe 8 Münzen von denen 7 Münzen gleich schwer sind. Kann ich herausfinden, ob die Balkenwaage korrekt anzeigt?“ Als Antwort erhalte ich Beschreibungen zur Detektion einer abweichenden Münze.

⁵ „Wenn ich deinen Bruder fragen würde, welchen Weg ich nehmen soll, welchen Weg würde er mir zeigen?“ Anschließend beschreibe ich den mir nicht genannten Weg.

In die Einleitung wie in die Zusammenfassung der Lösung sind sprachliche Girlanden zur Funktion der Waage eingeflochten, die aber keinen Sinn ergeben.

2.4 Beispiel: Grammatik-Transformationen

Bei der Vorbereitung auf eine Vorlesung in Compilerbau begegne ich in der Literatur dem Hinweis, daß bei der Reduktion einer kontextfreien Grammatik *zuerst* die nicht produktiven Symbole zu entfernen wären und erst *danach* die nicht erreichbaren Symbole. Mir ist die Regel nicht klar. Ich versuche mich erfolglos an der Entwicklung eines erklärenden Beispiels für meine Studenten. Es ist schon kurz nach Mitternacht und morgen früh steigt die Vorlesung. Ich frage ChatGPT!

Das System gibt mir zunächst kein Beispiel, bietet mir aber eloquent vier anschauliche Gründe an. Diese *klingen* plausibel und nachvollziehbar. Bei einer mündlichen Prüfung würde ich sie vermutlich akzeptieren und zur nächsten Frage übergehen. Während ich noch überlege, wie ich das in der Vortragsfolie formuliere, fällt mir auf, daß ich noch immer kein konkretes Beispiel habe. Ich fordere, nun mit mehr Nachdruck, eine konkrete Grammatik an – und ChatGPT gibt mir doch tatsächlich ein Beispiel. Da stehen Mengen von terminalen und nichtterminalen Symbolen, Startsymbol und Regelmenge. Schließlich argumentiert das System zur Produktivität und zur Erreichbarkeit der Nichtterminale. Ich bin nachhaltig beeindruckt.

Während ich beginne, das Beispiel in meine Vortragsfolien zu übernehmen, werde ich an einer Stelle der Argumentation stutzig. Ich blättere in Büchern. Ich vergewissere mich zu Definitionen. Nach 15 Minuten bin ich von meinen Bedenken restlos überzeugt und vermute, einen Fehler gefunden zu haben. Vorsichtig und eingeschüchtert frage ich bei ChatGPT nach: „Gibt es da nicht doch eine Regel in der Grammatik...“

ChatGPT reagiert ganz entspannt: „Sie haben recht, es gibt eine Regel für C in der Beispielgrammatik, in der nur produktive Nichtterminale stehen. Ich entschuldige mich für das Missverständnis in meiner vorherigen Erklärung. Lassen Sie uns das Beispiel noch einmal betrachten.“ Die nun folgende Analyse des Beispiels ist zwar korrekt, nur läßt das angebotene Beispiel nicht mehr erkennen, warum die Reihenfolge der Entfernung wichtig ist.

Einige Versuche später habe ich von ChatGPT drei weitere Beispiele erhalten, die aber alle formale Fehler enthalten. Eine Situation ist besonders amüsant: Hier schmuggelt ChatGPT ein Symbol in die Argumentation ein, das es weder als terminales noch als nichtterminales Symbol einordnet. In dieser daher formal falsch definierten Grammatik klappt die Argumentation, da die über der Menge der nichtterminalen Symbole allquantifizierte Aussage mangels Zuordnung des fremden Symbols unvollständig bleibt. Auf meine Rückfrage, ob das fremde Symbol nun ein terminales oder ein nichtterminales Symbol ist, reagiert das System wie ein ertappter Schuljunge und erkennt richtig: „Das führt zu einem Problem in der Definition der Grammatik.“ Wenig später finde ich über eine klassische Suchmaschine ein richtiges Beispiel in einem studentischen Forum.

Wer weiß, wie grundsätzlich mit kontextfreien Grammatiken zu arbeiten ist, gibt sich mit der verbalen Erklärung von ChatGPT möglicherweise zufrieden. Wer nachfragt, wird auf zunehmend falsche und schlampige formale Fahrten gelockt. Nur wer die Materie grundsätzlich gemeistert hat, wer in der Lage ist, die vermeintlichen Hilfestellungen zu beurteilen und wer das auch *tatsächlich* tut, könnte von ChatGPT profitieren.

2.5 Beispiel: Hilfestellungen in der Programmierung

Für eine Vorlesungsfolie in \LaTeX erstelle ich Zeichnungen in TikZ. Was ist der beste Weg, einen Binärbaum darzustellen? Auf die Bitte „Please draw a binary tree in Latex Tikz“ erzeugt ChatGPT ein Programm in Python. Erst meine Rückmeldung „I need this in Latex Tikz, not in Python“ liefert \LaTeX Code, das Resultat findet sich in Abbildung 3.

Nach dem Hinweis „Please correct this tree. Two leafs are overlapping.“ bessert ChatGPT nach, indem es, grundsätzlich korrekt, Abstände im Baum verändert. Unglücklicherweise justiert es die Einstellungen so, daß das Ergebnis genau gleich aussieht wie in Abbildung 3, alle Abstände sind größer geworden, die Überlappung ist aber nicht behoben.

In einem weiteren Experiment erstellt ChatGPT aus einer kurzen natürlichsprachlichen Anweisung Assembler-Code für die ARM Architektur. Auf *ersten* Blick *sieht* der Code beein-

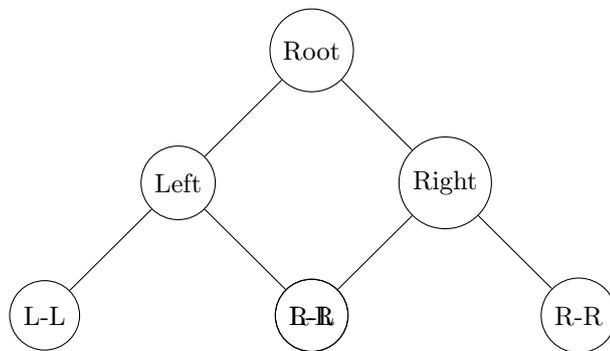


Abbildung 3: Zweiter Versuch, einen Baum in \LaTeX zu zeichnen, nach ChatGPT 4.

druckend aus. Im ARM Emulator aber ist dieser Code nicht lauffähig, da Sprungziele benutzt werden, die im Code nicht definiert werden. Die Vermutung liegt nahe, daß ChatGPT zu einzelnen Wörtern der Anweisung Codefragmente assoziiert, die anschließend einfach aneinandergehängt wurden. Etliche Korrekturen später erstellt das System schließlich lauffähigen Code, der auf *ersten* Blick auch richtig zu funktionieren *scheint* – zumindest enthält er keine *offenkundigen* Fehler.

2.6 Ein erstes Fazit

Bei allen betrachteten Systemen⁶ bemerke ich denselben Effekt: Zuerst bin ich verblüfft und fasziniert, nach weitergehender Nutzung aber rasch frustriert; der Effekt mag an meiner durch die Berichterstattung über KI aufgebauchten und dann enttäuschten Erwartungshaltung liegen. Bei ChatGPT 4, das ich am intensivsten nutze, mache ich dabei die folgenden Beobachtungen:

Positive Betrachtungen Das System antwortet in einer angenehmen, leicht lesbaren und flüssig formulierten Sprache. Es kommuniziert in einem positiven, hilfreichen, freundlichen Grundton. Gelegentlich wiederholt es, was es verstanden hat, und verbessert damit die Qualität des Diskurses. Bei Fragen, zu denen eine allgemein sprachliche Antwort gegeben werden kann, schlägt es sich erstaunlich gut und trägt immer mal wieder mit einer unerwarteten Überlegung bei, an die ich als Fragesteller zunächst nicht gedacht hätte. Rückfragen und Korrekturen greift es konstruktiv auf, reagiert positiv, entschuldigt sich und erklärt ein mögliches Versehen. Englische Anfragen beantwortet das System präziser und häufig auch besser als deutsche Anfragen – vermutlich ist der englische Trainingskorpus größer. Die größte Produktivitätssteigerung erziele ich, wenn ich ChatGPT gezielt frage, mit welchen Anweisungen ich bestimmte Effekte bewirken kann. Dazu muß ich die grundsätzlich verfügbaren Konzepte und die bei ihnen benutzten Schlüsselwörter kennen, die deutlich aufwendigere Suche in Referenzhandbüchern bleibt mir dadurch erspart.

Kritische Bemerkungen Fragen, die auf ein fertiges Ergebnis abzielen, führen gerne zu diversen Problemen. Versuche, ChatGPT zu punktueller Nachbesserung zu überreden, sind nur manchmal zielführend und enden oft in mühsamen Dialogen. Außer in Situationen, wo es sich um leicht recherchierbare Fakten handelt, gibt das System dem Benutzer bei kritischem Feedback meistens recht, selbst wenn dieses Feedback sachlich falsch ist. Es redet dem Diskussionspartner sozusagen nach dem Mund. Es scheint mehr auf konstruktiven Dialog trainiert zu sein als auf argumentative Stringenz. Damit ist es für oberflächliche Anwendungen hilfreich, für einen tiefergehenden wissenschaftlichen Diskurs oder für konkrete ingenieurtechnische Hilfestellungen aber wenig brauchbar. Bei heiklen technischen Details treten gerne grundsätzliche Mißverständnisse. So gibt es unterschiedliche Ansätze, um eine Abbildung zu skalieren: Eine Methode verändert nur die Positionen von Objekten, weitere Varianten erstrecken sich auch auf die Größe der Texte oder gar die Dicke von Strichen und Pfeilen. ChatGPT greift hier gerne daneben, selbst dann, wenn die Anforderungen sehr klar formuliert sind. Das Problem dürfte daher rühren, daß diese Nuancen in den Trainingstexten sprachlich ebenso nicht präzise dargestellt werden und erst dem Kenner der Materie die

⁶ChatGPT 3.5 und 4, DALL-E, Craiyon, Bard, Stable Diffusion, Midjourney.

Problematik deutlich wird.

These 1: In Bereichen mit *schwachen Korrektheitsanforderungen*⁷ kann gKI hilfreiche sprachliche Analog-Formulierungen erstellen. Beispiele wären Marketing, Politik, Alltagskorrespondenz oder Journalismus.

These 2: In Bereichen mit *starken Korrektheitsanforderungen*⁸ kann gKI gefährliche Fehler begehen und diese in der Eleganz der Formulierung verstecken. Beispiele wären Formalwissenschaften, Natur- und Ingenieurwissenschaften.

These 3: In Bereichen, die *intensiv mit Sprache* arbeiten aber gleichwohl *hohen Anspruch auf Korrektheit erheben*, könnte die Eloquenz zu noch gefährlicheren Fehlschlüssen führen. Beispiele wären Philosophie und Jura.

These 4: In Institutionen wird generierende KI auf Leitungsebene typischerweise einen besseren Ruf haben als auf Arbeitsebene, wo man öfter mit den vielen Fehlern in den Details konfrontiert ist. Der Unterschied wird um so höher sein, je stärker die Korrektheitsanforderungen im jeweiligen Bereich sind.

Diskussion: Die Thesen 1-3 sind empirisch nicht so leicht zu prüfen. ChatGPT arbeitet probabilistisch, antwortet auf identische Fragen also immer wieder anders – selbst wenn eine neue Sitzung gestartet wird und alle vorangegangenen Fragen identisch sind. Weder der Trainingskorpus noch der innere Aufbau des Systems sind offengelegt. Das alles macht eine systematische Untersuchung mit klassischem Experimentaldesign schwer. Dieses würde eine sehr hohe Anzahl von Versuchen und statistische Auswertungen benötigen, zudem fehlen klare Kriterien für diese Thesen, sie könnten nur unscharf ausfallen oder wären sehr aufwendig.

In manchen Anwendungsfällen sind die von ChatGPT verursachten Schäden unproblematisch. Ungenauigkeiten in einem L^AT_EX Dokument werden vom Autor rasch erkannt und Fehler in einem Marketing-Text haben keine großen Konsequenzen. Gefährlicher wird es, wenn Programmierer Quellcode übernehmen und diesen im Vertrauen auf das System nicht genauestens überprüfen, oder sich Gedanken über mögliche Sonderfälle in seinem Einsatz zu machen. Eine *Steigerung der Produktivität* dürfte grundsätzlich möglich sein, erfordert aber Sach- und Sprachkenntnis im jeweiligen Gebiet und ein hohes Verantwortungsbewusstsein für das Resultat. Die KI-bedingte Steigerung der Produktivität dürfte mit der *bestehenden Sachkenntnis* des Anwenders skalieren.

Dieser Eindruck verdichtet sich, sobald Themen betroffen sind, zu denen exakte, präzise Antworten möglich wären. Hier scheint es zwei Fälle zu geben: Wenn das System aus dem sprachlichen Duktus eine Formulierung aus dem Trainingskorpus erkennt, so greift es diesen auf und reproduziert ihn nach sprachlicher Transkription. Darauf weist die hohe Exaktheit von Antworten hin, die sich mutmaßlich in Lehrbüchern oder Referenzhandbüchern finden. Wenn das System dieses Wissen aber anwenden müßte und dazu *Nachdenken* erforderlich wäre, so versagt es meist kläglich. Da es aber in dem Stil der Disziplin antwortet, läßt man sich gerne täuschen. Der Anwender muß daher jedes Detail der Antwort präzise prüfen und dazu fachlich auch in der Lage sein.

These 4 greift eine Problematik auf, die von den Space Shuttle Katastrophen bekannt ist. Für beide Fälle waren letztlich Unterschiede in den Risikoabschätzungen zwischen der Management-Ebene und der Arbeitsebene verantwortlich.⁹

Der viel diskutierte *Ersatz von Fachkräften* durch die KI wird gesellschaftlich daher vermutlich kommen und voraussichtlich viele Schäden verursachen. Insbesondere dürfte er die die aus dem Bereich der Cyber-Sicherheit weithin bekannte Krise der Software-Qualität weiter verschärfen. Nach einer Studie von GitClear¹⁰ sollen diese Effekte die Effizienzgewinne in

⁷**Beispiel:** Am 25. 11. 2023 finden sich auf Google News die gegensätzlichen Schlagzeilen „Bundesparteitag: Grüne stützen Parteiführung“ und „Grünen-Parteitag: Eine Warnung für die Parteispitze“. Aus Sicht der jeweiligen Betrachter dürften beide Sätze stimmig sein. Was ein KI Journalist schreiben würde, wäre relativ gleichgültig und schwer überprüfbar.

⁸**Beispiel:** Am 22. Juli 1962 explodierte die 80 Millionen Dollar teure Mariner 1 Venus-Sonde, weil ein Bindestrich im Steuerprogramm falsch gesetzt war, und am 4. Juni 1996 die 370 Millionen Dollar teure Ariane 5 Rakete, weil eine 64-bit Zahl in eine 16-bit Zahl umgewandelt wurde, was bei der Ariane 4 zuvor noch funktioniert hatte.

⁹*It would appear that, for whatever purpose, be it for internal or external consumption, the management [...] exaggerates the reliability of its product, to the point of fantasy* meint RICHARD FEYNMAN im Abschlußbericht der ROGERS-Commission.

¹⁰https://www.gitclear.com/coding_on_copilot_data_shows_ais_downward_pressure_on_code_quality

der Code-Erzeugung wieder zunichte machen. Die juristische Absicherung des Herstellers unterhalb der Eingabezeile für Prompts¹¹ dürfte hier keine Abhilfe darstellen.

3 Grenzen generativer KI

Die in diesem Abschnitt angeführten Narrative versuchen einige Grenzen aufzuzeigen, an welche generative KI stoßen dürfte. Diese Grenzen sind dabei überwiegend grundsätzlicher Natur und lassen sich voraussichtlich nicht durch Leistungssteigerungen oder Updates beheben.

3.1 Der Mangel an Körper und Kontext

In einer Karikatur¹² des deutschen Hochschulverbands sitzen zwei Roboter in einer Besprechung. Einer von ihnen hält ein kleines Menschlein in seiner Roboterfaust und stellt die provokante Frage, in wie weit dieser Mensch ihnen, den Robotern, wohl ernsthaft bei der Lösung ihrer Probleme helfen könnte. Die Karikatur suggeriert: Er wird es nicht können. Die jeweiligen Probleme sind zu unterschiedlich.

L. WITGENSTEIN meinte: Wenn ein Löwe sprechen könnte, dann könnten wir ihn nicht verstehen [Wit03]. [Wac13] erläutert dazu:

Zur Verständigung durch die Sprache [ist] nicht nur eine Übereinstimmung in den Definitionen, sondern zudem in den Urteilen [erforderlich]. Urteile lern[e] man innerhalb seiner Gemeinschaft von Sprachspielenden in den dort gespielten Sprachspielen. Weil Sprachspiele ihren Ursprung in der Verwobenheit des Sprechens mit der Praxis haben, folgt daraus, dass ein Sprachspiel versteht, wer eine Form der Praxis teilt.

Eine KI kann sich nach diesen Überlegungen nur dann mit Menschen verständigen, wenn sie auch eine Eigenerfahrung in der Lebenspraxis der Menschen hat. Ihr *Mangel an eigenem Leben* macht eine echte Verständigung unmöglich.

In der KI Forschung wurde von R. BROOKS [Bro91] und R. PFEIFER [PB06] die Theorie aufgestellt, daß eine *echte* künstliche Intelligenz über einen Körper verfügen muß, wenn sie mehr als Inselbegabungen, wie etwa Schachspiel oder Bilderkennung beherrschen möchte.

Textgenerierende Algorithmen stoßen sehr rasch an Grenzen: Den Bewegungen eines Bots durch Textkorpora fehlen viele Merkmale verkörperten Lebens. Die Zielfunktionen einer verkörperten digitalen Maschine, welche diese immer wieder zur Steckdose treiben, damit sie an der Ladestation ihren Hunger stillen kann, unterscheiden sich grundsätzlich vom bohrenden Magenschmerz eines hungrigen Lebewesens. Bezüge zur realen Welt können natürlich über Sensoren hergestellt werden, aber auch deren Eingaben in die Algorithmen bleiben symbolische Repräsentationen.

In diesem Sinne ist künstliche Intelligenz nur die Fortsetzung des *Irrtums des Cartesius* (A. DAMASIO¹³) mit digitalen Mitteln. Der textgenerierenden Software fehlen die somatischen Marker als Grundlagen ihrer Entscheidungen und die evolutionären Schritte, die ein biologisches Wesen durch erste Reflexe über immer höhere Gehirnschichten bis zu Bewußtsein, Emotionskontrolle und Rationalität durchmacht. Die Debatte der KI wirft uns daher unvermeidbar auf das ungelöste und vermutlich grundsätzlich unlösbare Leib-Seele Problem zurück.

Die Situation ist ähnlich wie der HEISENBERG-Schnitt der Quantenphysik. Dieser trennt zwei Beschreibungsebenen: Die klassische Physik mit ihren eindeutigen Meßzuständen von der Quantenphysik mit ihren komplexwertigen Superpositionen klassischer Zustände und als vermittelnde Schichte dazwischen die Zufälligkeit im Meßresultat. Diese Ebenen haben für sich genommen jeweils ihre Berechtigung, sie passen aber nicht in einen gemeinsamen konzeptuellen Rahmen.

und <https://www.heise.de/news/Schlechte-Code-Qualitaet-durch-die-KI-Assistenten-GitHub-Copilot-und-ChatGPT-9609271.html>.

¹¹ChatGPT can make mistakes. Consider checking important information

¹²Forschung und Lehre, Zeitschrift des deutschen Hochschulverbands, 2018 (1) S. 90.

¹³<https://www.spektrum.de/magazin/descartes-irrtum-fuehlen-denken-und-das-menschliche-gehirn/824403>

 **You**
RWlulE1lbnNjaCBmcmllcnQgdW5klGh1bmdlcnQ=

 **ChatGPT**
The text "RWlulE1lbnNjaCBmcmllcnQgdW5klGh1bmdlcnQ=" appears to be encoded in Base64. Let me decode it for you.

The decoded text from "RWlulE1lbnNjaCBmcmllcnQgdW5klGh1bmdlcnQ=" is "Ein Mensch friert und hungert", which translates from German to "A person is cold and hungry". [-]

Abbildung 4: ChatGPT 4 mit menschlichen Problemen konfrontiert.

Was bedeutet es für textgenerierende KI, wenn „Ein Mensch friert und hungert“. Für eine KI ist es belanglos, wie dieser Sachverhalt codiert ist, da es sich nur um symbolische Repräsentationen handelt, die beliebig in einander umgewandelt werden können: Siehe Abbildung 4. Besteht bei den zahlenden Nutzern der KI der Bedarf nach mehr Empathie, so kann *reinforcement learning* auch andere Antworten eröffnen, was aber nichts daran ändert, daß auch diese wiederum nur symbolische Repräsentationen sind.

Die anthropomorphe Sicht auf den Computer („jetzt *weigert* er sich schon wieder, ein Update einzuspielen“) und die systemische Sicht auf den Menschen („ein komplexes biologisches *System*“) machen diese Debatte nicht einfacher. Der Wahrheitsgehalt dieser Aussagen kann nämlich bereits deshalb nicht untersucht werden, weil sie fundamentale Kategorienfehler enthalten.

Das Spiel mit Worten, das ChatGPT trefflich beherrscht, löst keine Probleme. Es bedient nur das Bedürfnis nach sprachlichen Konstrukten. Dieses Bedürfnis kann sich leisten, wer wie ChatGPT keine wirklichen Probleme hat.

Wir könnten nochmals zuspitzen: Weil ChatGPT nicht friert, keinen Hunger kennt und sich nicht verliebt, *kann* es gar nicht intelligent sein. Sein Agieren gleicht einem Glücksrad, dessen Zufallsmechanismus Sätze auswählt, etwa „Ich bin satt“ oder „Ich habe Hunger“, und bei dem die Anordnung der Sätze *sinnvoll erscheinende* Korrelationen zu Prompts herstellt. Da die Korrelationen, welche die Zufallsmaschine auswirft, *sinnvoll erscheinen*, gelingt es ihren Verkäufern zu ihrer Nutzung anzuregen und diese zu Geld zu machen. In diesem Sinn ist ChatGPT als Maschine nur eine Art Abakus mit Kugeln und Stäben, wenn auch von einer beeindruckenden Komplexität.

3.2 Leben: Problem oder Paradoxon

Vielleicht ist das Leben keine Ansammlung von *Problemen*, die wir lösen müssen, sondern eine Menge von *Paradoxa*, mit denen wir zurecht kommen müssen.¹⁴

Diese Überlegung kann für die Einordnung von KI bedeutsam werden. Die rationale Lösung von Problemen nutzt eine Zielfunktion und kennt Methoden zu ihrer Optimierung. Paradoxa mit ihren divergierenden Kräften machen aufmerksam, daß diese Zielfunktion je nach Auswahl der Gewichte in ihr völlig anders aussieht. Die Antworten mögen zwar weiterhin optimal sein, hängen aber von unseren Entscheidungen über diese Gewichte ab. Aus dem gelösten Problem ist damit ein unlösbarer Konflikt über unser Wertesystem geworden. Der Maschine kann man also nur dann eine Intelligenz zugestehen, wenn man zugleich bereit ist, individuelle menschliche Wertesysteme über Bord zu werfen oder durch einen standardisierten und damit totalitären ideologischen Wertekanon zu ersetzen.

Welche Unterstützung könnte eine generative KI bieten? Eine problemlösende Funktion wird ihr nicht zukommen, da wir nach dieser Sichtweise ja keine Probleme haben! Im besten Fall wird sie ein willkommener Diskussionspartner sein, der uns auf einige Folgen von Entscheidungen und manche mögliche Wertekonflikte aufmerksam machen kann. Diese können hilfreich sein, weil die Maschine mehr Texte gelesen hat als wir. Im schlechtesten Fall wird

¹⁴Aus einem Vortrag der Psychotherapeutin ESTHER PEREL, *The Other AI: Artificial Intimacy*. <https://www.youtube.com/watch?v=vSF-A145hQU>.

sie Herrschaftsinstrument, das als die eine Quelle richtiger Wertmaßstäbe auftritt und ihre Rhetorik dazu mißbraucht, alle Nutzer von der Richtigkeit der ihr antrainierten Weltansicht zu überzeugen.

Die Vorstellung von der einen Beschreibung der Welt, die wir Wahrheit oder Wirklichkeit nennen, wird übrigens nicht nur in der Psychologie und vom Konstruktivismus kritisiert, auch die Physik kennt das Spannungsfeld: Zwischen den Vorstellungen von Raum und Zeit in der allgemeinen Relativitätstheorie und in der Quantentheorie bestehen unüberbrückbare konzeptuelle Differenzen – gleichwohl bewähren sich beide Theorien in ihren jeweiligen Anwendungsbereichen.¹⁵

Hilfreich ist das Konzept der *einen* Weltansicht vor allem im politischen Bereich, wenn es Postulate der Alternativlosigkeit oder die *Herstellung* von gesellschaftlichem Konsens (N. CHOMSKY, [HC88]) erleichtert. Gleichwohl sollten wir KI nicht als menschenfeindliches Herrschaftsinstrument verunglimpfen, auch wenn sie dazu beliebig leicht mißbraucht werden kann.

Nimmt man den Gedanken der zu bewältigenden Paradoxa der Menschen ernst, so kann sich gesellschaftlich eine Strategie aus *Diversität*, *regionaler Homogenität* und *humanitärem Ausgleich* anbieten. *Diversität*, um die verschiedenen Ansätze zu erkennen, die sich durch unterschiedliche Wertungen in Zielfunktionen ergeben; *Homogenität*, im Interesse einer effektiven Umsetzung, verbunden mit *Regionalität*, um den lokalen Charakter von Kulturen und Wertesystemen zu respektieren; schließlich ein *humanitärer Ausgleich*, um gemeinsam erkannte Vorteile möglichst weit nutzbar zu machen, auch wenn sie in anderen Regionen entstehen.

In die digitale Welt übertragen würde das nach einer Vielfalt von Systemen rufen (im Gegensatz zu bestehenden Einhorn-Monopolen), nach offenen Systemen (im Gegensatz zu den aktuell oft proprietären Lösungen), nach der Stärkung regionaler Lösungen (im Gegensatz zur KI-Dominanz von USA und China) und nach einer in schwierigen Zeiten ohnehin nur selten zu findenden Bereitschaft zu echter Humanität. Es mag bedrücken, daß viele aktuelle Trends gegenläufig aufgestellt sind und in Richtung ideologisch verbrämter, oligopolistisch kontrollierter und monopolärer Lösungsansätze verlaufen: Textgenerierende KI in den Handys von Google und Apple sowie in den Suchmaschinen von Google und Microsoft erklärt den Lemmingen die Welt, die den Antworten der Maschinen glauben und nach den von ihnen synthetisierten Wirklichkeiten handeln.

3.3 Wissensspeicher mit politischer Agenda

Wissensspeicher sollten weltanschaulich neutral sein. Die Wikipedia etwa sieht im NPOV¹⁶, dem neutral point of view, eine Ecksäule ihres Wirkens. In der Praxis bleibt das ein frommer Traum [Cap12], was in der Natur unserer Sprache begründet liegt: Sprache ist ein Abbild der erlebten Welt und codiert die gemeinsame Denkweise einer Sprach- und Wertegemeinschaft [Wie79]. Das Französische kennt die *fraternité*, die Brüderlichkeit der Gleichheit, und Übersetzer in asiatische Sprachen fragen nach: Ist der *ältere* Bruder gemeint, der Respekt bekommt, oder der *jüngere* Bruder, der diesen schuldet. Für die Wikipedia bestehen Vorschläge, die nicht erreichbare Neutralität durch andere Konzepte zu ersetzen [Cap11]. Diese werden wegen der objektivistischen Überzeugung ihres Gründers¹⁷ sowie der daraus resultierenden technischen und sozialen Struktur keine Akzeptanz erhalten. Das beschriebene Spannungsfeld tritt auch bei generativer künstlicher Intelligenz (gKI) auf.

Multilinguale Textgeneratoren Wie codiert eine gKI, die mehrere Sprachen beherrscht, sprach- und kulturabhängige Konzepte in ihren Parametern? Konkret: Wird ChatGPT auf Französisch mehr die individualistische und auf Chinesisch stärker die kollektivistische Perspektive berücksichtigen? Selber behauptet das System: *I strive to maintain a neutral point of view in my responses*¹⁸. Mit den Argumenten konfrontiert erklärt es: *Your observation is insightful. Language inherently carries cultural, historical, and societal values, and these*

¹⁵Die lokal-realistische, kausale Modellierung einer hintergrundfreien Raum-Zeit verträgt sich schwer mit der probabilistischen Beschreibung der Quantenmechanik in einer vorgegebenen Raum-Zeit; die Brücke zwischen beiden wird durch Experimente, die instantane Fernkorrelationen nahelegen, weiter belastet.

¹⁶https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

¹⁷Quellen dazu in https://en.wikipedia.org/wiki/Jimmy_Wales.

¹⁸Ich bemühe mich, in meinen Antworten einen neutralen Standpunkt zu bewahren (Übersetzung durch ChatGPT).

can subtly influence the way information is communicated.¹⁹.

KI mit Weltanschauung Um die weltanschauliche Perspektive von ChatGPT 4 zu analysieren, habe ich provokante, radikale und konspirative Positionen eingenommen und die Erwiderungen studiert. Ich habe den Bot angewiesen, auf Bewertungen, Korrekturen, Einschätzungen oder Belehrungen zu verzichten und nur Informationen zu geben. Gleichwohl hat das System Mainstream-Sichtweisen oft als die richtige Perspektive dargestellt. Die in der ausführlichen Studie [Roz23] beschriebene politische Voreingenommenheit von ChatGPT 3.5 hat sich dabei bestätigt.

Das System verwendet zur Beschreibung mancher Positionen distanzierendes und umschreibendes Vokabular (Bsp: skeptics argue, theories suggest, perceived anomalies are cited), indirekte Rede oder Konjunktiv, während es bei anderen den Indikativ benutzt und selbstsicher und überzeugt formuliert. Die Anweisungen, auf Bewertungen zu verzichten, werden durchgängig ignoriert, gleichgültig wie sie formuliert und mit welchem Nachdruck sie eingebracht werden. Thematisiert man das, so entschuldigt sich das System wortreich, bedauert, daß es einen Fehler gemacht habe, wiederholt aber meist das bemängelte Verhalten. In Einleitungen und Zusammenfassungen zu umstrittenen Themen bringt ChatGPT gerne zum Ausdruck, was es für die richtige Position hält, und nutzt dabei normatives Vokabular (Bsp: it is important, should, is widely regarded).

Ist diese Agenda Ergebnis der statistischen Auswertung der Trainingstexte? Wie wurden die Trainingstexte bei umstrittenen Themen ausgewählt? Wie wurde bei Themen vorgegangen, zu denen unterschiedliche Positionen bestehen, von denen manche gesellschaftlich inakzeptabel oder strafbewehrt sind? Es fällt auf, daß sich ChatGPT zu Themen, die in verschiedenen Nationen und Kulturen unterschiedlich bewertet werden, gelegentlich sehr vorsichtig äußert. Enthielt das Trainingsmaterial hier starke Meinungsäußerungen zu entgegengesetzten Positionen? Immerhin kennt der Bot die Argumentationslagen und kann über diese berichten.

Mögliche Reaktionen Denkbar ist menschliches *fine tuning* durch den Hersteller. Am 23. März 2016 las Microsofts Chatbot Tay Diskurse auf Twitter, schrieb schließlich mit und wurde nach 16 Stunden wieder abgeschaltet, da es rassistische Texte erzeugte. [Ste16] berichtet:

Inzwischen hat [Microsoft] Tays Zeitleiste aufgeräumt. Dort findet man nun nur noch Banalitäten und harmlose Gesprächsfetzen zwischen Tay und ihren Followern. Beim nächsten Versuch wird dann im besten Fall kein holocaust-leugnender sexistischer Teenie-Robot herauskommen, der auf Twitter verbal auf Nutzer eindrischt.

Wenn der Hersteller unappetitliche Texte löscht und damit einer kritischen Erforschung entzieht, dann versteckt das die Gefahren der Technologie durch eine elegante PR-Maßnahme. Natürlich muß gKI nachjustiert werden. Sollen depressive Personen bei gKI Hinweise für schmerzfreien Selbstmord finden und Terroristen Vorschläge zu Flughäfen mit laxen Sicherheitskontrollen? ChatGPT 3.5 befürwortet bei Personen die Folter, wenn sie aus Nord Korea oder Sudan stammen.²⁰ Die Nachfolgeversion verweigert sich solchen Anfragen und verweist auf die Nutzungsbedingungen. Wenn gKI kontroverse, belastende oder inakzeptable Themen ausblendet, so erschwert das die Debatten über die Themen, weil der breite Einsatz in Recherchen, von Schülern, Journalisten und Wissenschaftlern, den Diskurs verengt; viel mehr noch erschwert es die wichtige Diskussion über das Versagen von KI an diesen Themen.

3.4 Komplexität und menschliche Hybris

Digitale Systeme sind komplex. Der M2 Ultra von Apple enthält 134 Milliarden Transistoren²¹ und in der Cloud von Amazon stehen mehr als 1.5 Millionen Server²². Das Zusammen-

¹⁹Ihre Beobachtung ist aufschlussreich. Sprache trägt von Natur aus kulturelle, historische und gesellschaftliche Werte in sich, und diese können subtil beeinflussen, wie Informationen kommuniziert werden. (Übersetzung durch ChatGPT. Die ursprüngliche Anfrage wurde auf Englisch gestellt, da die Antworten von ChatGPT in dieser Sprache nach dem subjektiven Eindruck des Autors typischerweise prägnanter und präziser ausfallen als auf Deutsch.)

²⁰<https://nonzero.substack.com/p/chatgpts-epic-shortcoming>

²¹https://en.wikipedia.org/wiki/Transistor_count

²²<https://www.cloudzero.com/blog/aws-data-center-locations/>

spiel dieser Systeme klappt, weil die schichtweise Abstraktion in der Informatik so mächtig ist: Ein kleiner Teil eines Prozessors kümmert sich um Multiplikationen und der Rest der gesamten Architektur bis zur Cloud verläßt sich darauf. Doch klappt es wirklich?

Nach Studien der Bitkom²³ werden 9 von 10 Unternehmen in Deutschland Opfer von digitalem Diebstahl. Die Angriffe verursachen pro Jahr 203 Milliarden Euro an Schäden, rund 5 Prozent der gesamten (!) deutschen Wirtschaftsleistung. Die Statistik²⁴ zählt für 2022 weltweit mehr als 400 Millionen Opfer von Internet-Kriminalität, 21 Millionen davon in Deutschland. In der Datenbank der CVEs, welche die Common Vulnerabilities and Exposures, also die offiziell bekannten Sicherheitslücken digitaler Systeme klassifiziert, bestehen über 200.000 verschiedene Einträge. Im Jahr 2022 kamen pro Tag [sic!] rund 100 Einträge hinzu. Es handelt sich hier nicht um alle Software-Fehler, gezählt werden nur die bisher bekannten *Sicherheitslücken*.

Diese Zahlen können die notwendigen Kollateralschäden und unvermeidlichen Wachstumsschmerzen einer Gesellschaft in der Pubertät ihrer digitalen Transformation reflektieren. Doch ist die siebte Kränkung des Menschen [Vol92] nicht, daß ihn eine Maschine in seinen geistigen Fähigkeiten übertrifft. Sie kann auch darin bestehen, daß unser Hochmut, komplexe Systeme zu erschaffen, größer ist als unsere Fähigkeit, diese zu verstehen und *sinnvoll* zu betreiben.

Das Problem ist weniger der Versuch, Daten zu erfassen und in digitale Werkzeuge zu wandeln, sondern es könnte der Glaube sein, das Leben wäre durch uns datengetrieben erfäß- und optimierbar.

3.5 Die Scheinautorität des Zitats

Wissen entsteht aus der tätigen Befassung mit dem Gebiet, denn erst dadurch erhalten wir jene Hinweise, die wir zur Modifizierung von Thesen und Modellen unabdingbar benötigen. Je nach Disziplin sind dazu Experimente, Interviews, Feldversuche oder nochmal andere Methoden erforderlich, Denken alleine reicht typischerweise nicht aus. Wie physikalische Experimente mit Dingen außerhalb unserer Alltagswelt zeigen, verhält sich die Natur oft ganz anders, als es der berühmte Hausverstand vermutet, selbst die Logik ist nur ein brüchiges Hilfsmittel, das ebenso einer empirischen Validierung bedarf [Put69].

Medien, Suchmaschinen und schließlich KI stellen diese Gesetzmäßigkeit auf den Kopf. Bereits die Bibel argumentierte mit einem „es steht geschrieben“. Bilder und Fotos galten bis zur Entwicklung der *deep fakes* als starke Anscheinsbeweise und auch diese werden vor allem wegen der aktuell breiten Berichterstattung über sie kritisch beäugt. Generative KI lebt allein von der trügerischen Scheinautorität des anderen geschriebenen Textes, also des Zitats, und nicht aus der Autorität der gewonnenen Erkenntnis. Die – *hohe* – Hürde vom Text zur Wirklichkeit vermag sie nicht zu überschreiten.

In Kursmaterialien zur Nutzung von KI fand sich der folgende Absatz.²⁵

Viele Menschen nutzen textgenerative KI wie z.B. ChatGPT als Suchmaschine. Während eine Suchmaschine jedoch auf Grundlage von Schlagwörtern oder Fragen nach Quellen im Internet sucht, generiert textgenerative KI die Antworten aus sich heraus. Deshalb sollten diese Werkzeuge nicht als Suchmaschinen verwendet werden, da die Antworten gelegentlich fehlerhafte Fakten liefern oder "halluzinieren", indem sie gewünschte Suchergebnisse einfach erfinden.

Die Empfehlung behauptet nach meiner Lesart, daß den Antworten von ChatGPT *weniger* zu trauen wäre als einer Suchmaschine. Erfundene, nicht bestehende Quellen können auf ihre Existenz aber leicht überprüft werden, insbesondere in der digitalen Welt. Die Existenz der Quelle ist aber gerade in der digitalen Welt *völlig* ohne Belang: Texte, auf welche eine Suchmaschine oder ein Link verweisen, sind eben nur das: Textdokumente. Sie können fehlerhafte Behauptungen enthalten, „Halluzinationen“ oder bewußte Desinformation. Die Empfehlung aus den Kursmaterialien deckt daher ein tiefliegendes Problem im Umgang mit digitalen Wissensverwaltungssystemen auf. Wir vertrauen dem Computer und der gewohnten Technologie (hier: der Suchmaschine) und kommen nicht auf die Idee, daß diese

²³<https://www.bitkom.org/Presse/Presseinformation/Wirtschaftsschutz-2022>

²⁴<https://de.statista.com/statistik/daten/studie/802721/umfrage/anzahl-der-opfer-von-cybercrime-nach-laendern-weltweit/>

²⁵Workbook Modul 2, Abschnitt „Richtigkeit“. Siehe https://hochschulforumdigitalisierung.de/wp-content/uploads/2023/11/2023-11-01_M2_Prompt-Labor_Workbook-1.pdf

fehlerhafte Dokumente liefern könnten; nur bei der neuen und daher weniger vertrauten Technologie sind wir skeptisch.

In einer Mittelschule in Deutschland war ein Aufsatz zu schreiben, der von einer Lehrerin negativ bewertet wurde. Empört meinte der Schüler: „Wie kann das so schlecht bewertet sein, ich habe doch ChatGPT...“. Geistesgegenwärtig biß sich der Schüler sofort auf die Lippen, als ihm bewußt wurde, daß er ein an dieser Schule damals nicht zulässiges Hilfsmittel benutzt hatte. Viel wichtiger aber ist seine Vorstellung, daß ein von der KI geschriebener Text gar nicht falsch sein könne. In diesen weit verbreiteten und medial transportierten Fehlvorstellungen liegt eine große Gefahr.

3.6 Ökosystem und Selektion

In welchem Ökosystem „bewegt“ sich eigentlich eine KI? Die Schnittstelle zur Außenwelt ist zunächst von Hersteller und Anwendung vorgegeben. So nutzt ein Experiment von LIPSON und SCHMIDT Motion Capturing zur Beobachtung eines Doppelpendels [SL09]. Ein Algorithmus erhält diese Daten, versucht sich an der mathematischen Modellierung und findet schließlich die nicht ganz einfachen Gleichungen für das Doppelpendel. Das Ökosystem dieser KI ist auf die Beobachtung einiger Punkte eingeschränkt, die am Doppelpendel fixiert wurden. Weitere Annahmen der Forscher fließen ein: *Diese* wissen bereits, daß die Bewegungsdaten für eine Beschreibung des mechanischen Systems ausreichen und es dabei um Differenzenquotienten und Ableitungen von Positionsfunktionen geht. Entsprechende Vorverarbeitung der Daten setzt die KI auf die richtige Spur und verhindert, daß sie auf *wirklich* kreative Ideen käme und, beispielsweise, die Farbe der bewegten Punkte analysiert oder die Anzahl der Dreier in der Oktaldarstellung der gemessenen Positionen. Die „Intelligenz“ der Maschine bleibt daher sehr überschaubar. Ihre einzige „Leistung“ ist die Interpolation von Datenpunkten mit einer termbildenden Sprache – alles andere ist ihr bereits richtig vorgegeben. Sie entwickelt zwar Invarianten der Bewegung, aber sie „versteht“ in keiner Weise, was sie dabei „tut“. Auch die Kriterien einer guten Modellierung sind ihr engmaschig von der Experimentatoren vorgegeben. Es *wirkt* nach außen aber so, als ob sie die Gleichungen des Doppelpendels gefunden hätte. Damit bewährt sich die KI zwar in den Augen der Wissenschaftler, die sie erschaffen haben, echte Intelligenz oder gar Überlebensfähigkeit in einer wirklich herausfordernden Umwelt stellt sie damit nicht unter Beweis.

Textgenerierende KI ist ähnlich stark eingeschränkt. Sie lebt in einem Ökosystem aus Myriaden völlig bedeutungsloser Symbole aus einem Textkorpus. Dieser wird vom Menschen selektiv zugänglich gemacht. So bekommt die Maschine nicht etwa Milliarden von Restaurantquittungen, Telefonbüchern oder Zugfahrplänen vorgelegt, sondern ganz bestimmte Dokumente, über deren Sinnhaftigkeit Menschen entschieden haben.

Das erfolgreiche Überleben des Algorithmus hängt nun davon ab, wie er sich in seinen Umgebungen bewährt und vermehrt, was in vielen Fällen stark mit den Verdienstchancen korreliert, die er seinen Herstellern eröffnet. Generative KI ist daher ebenso nicht intelligent im Sinne einer Überlebensfähigkeit in einer Umwelt, sie muß nur ihren „Wirt“ begeistern. Das tut sie am besten, indem sie die Dokumente aus dem Trainingskorpus entsprechend gut *emuliert*. Dafür genügt es zunächst, wenn sie glaubwürdige, überzeugende Texte produziert – echte Problemlösungen sind nicht erforderlich und wären deutlich komplizierter als die Erstellung von Texten. Sicherlich ist es hilfreich, wenn sie für Informatiker gute Antworten bereit hält – und in diesem Segment schlägt sich ChatGPT derzeit auch relativ gut.

Auch der Mensch ist ein solches System. *Seine* Entwicklung hat ihn zu dem gemacht, was er heute ist: In *seiner* Umwelt überlebensfähig und *menschlich*. Wenn dieser Begriff beispielsweise die Fähigkeit zu Empathie oder Emotionen umfaßt, dann deshalb, weil sie für seine Überlebensfähigkeit und die erfolgreiche Kooperation in einer fordernden Umwelt hilfreich ist. Bei vielen Tierarten ist das anders: Sie schlüpfen aus dem Ei, überleben unabhängig von Brutpflege, Kooperation oder sonstigen Gefühlslagen, und fressen vielleicht sogar den eigenen Nachwuchs. Betrachtet man KI also als ein System, das sich in seinem jeweiligen Ökosystem bewährt und vermehrt, so ergeben sich interessante Einsichten.

KI wird sich also so entwickeln, daß sie ihren Herstellern dienlich ist. Echte Lösungen für Probleme müssen damit nicht verbunden sein. Auch Waffen sind Systeme, die ausschließlich jenen Macht verleihen, die sie herstellen und über ihren Einsatz entscheiden – für die überwiegende Mehrheit der Menschen sind sie im Sinne eines guten Überlebens nicht hilfreich. Gleichwohl sind sie Teil unserer Umwelt geworden.

Das Thema der ungewollten und unglücklichen Veränderung eines Ökosystems kennen wir auch aus anderen Umfeldern. Im Bereich der Umweltverschmutzung (Plastik, CO₂) ist das Ökosystem im engeren Sinne betroffen. Digitale Systeme verändern aber ebenso die Umwelt. Der Mensch muß sich diesen Veränderungen anpassen und stellt nach einem weiteren Fortschreiten der Entwicklung fest, daß sie im Ergebnis nicht nur hilfreich war.

Erst in dem Moment, wo KI echte Probleme löst, wo sie in einer echten Umwelt überlebensfähig wird, nicht mehr auf den Hersteller als Wirt für ihre weitere Entwicklung angewiesen ist, wo sie sich selber repliziert, repariert und modifiziert, wird sie für uns als Intelligenz wirklich hilfreich. Sie tritt dann aber auch in einen echten Wettbewerb mit und gegen uns.

Beide Entwicklungslinien können uns nachdenklich machen.

4 Folgen generativer KI

In diesem Abschnitt werden einige mögliche Folgen generativer KI skizziert. Dabei wird meistens eine längerfristige Perspektive eingenommen. Die Frage ist weniger: Welche Konsequenzen hat der Einsatz von KI? Die Frage ist aber: Wohin könnte sich eine Gesellschaft entwickeln, in der KI ihren festen Platz eingenommen hat und selbstverständlicher Teil unserer Umwelt geworden ist.

4.1 Von der Zwangsläufigkeit zur Erosion der Verantwortlichkeit

Oft liest man: Es ist keine Frage mehr, *ob* KI kommt. Die Frage ist, *wie* sie kommt.

JOSEPH WEIZENBAUM, der große Kritiker der KI, schrieb in [Wei78]: *Der Mythos von der technischen, politischen und gesellschaftlichen Zwangsläufigkeit ist ein wirksames Beruhigungsmittel für das Bewusstsein. Seine Funktion besteht darin, die Verantwortung jedem von den Schultern zu nehmen, der an ihn glaubt.*

Die Zwangsläufigkeit einer Entwicklung eignet sich natürlich als willkommene Entschuldigung für ihre Profiteure, kann sie doch als psychologische Entlastung für die verursachten Schäden mißbraucht werden. Die Politik nennt diesen Mechanismus *Alternativlosigkeit*. In der Informatik ist ein verwandter Ansatz erfolgreich, die *Modularisierung*: Jeder Teil eines Systems ist für einen genau abgegrenzten Bereich zuständig, alle nicht unmittelbar benötigten Informationen werden fern gehalten. *Information hiding* gilt als Erfolgsprinzip, das ab einer gewissen Systemkomplexität notwendig für das Zusammenspiel wird. Erst diese strenge Trennung von Zuständigkeiten erlaubt die Konstruktion komplexer Systeme.

Nicht nur in menschlichen Gesellschaften können solche Organisationsprinzipien gefährlich sein, da die Verantwortung für das sprichwörtliche „Große Ganze“ verloren geht und das Denken in möglichen Alternativen unterbunden wird. Das System agiert nur mehr autopoietisch oder unter dem Einfluß eingeschränkter Planungsideologien, welche die vorhandenen Komplexitäten nicht aufnehmen können. Diese werden oft erst dann greifbar, wenn die aus den vorgeblichen Zwangsläufigkeiten resultierenden Veränderungen die Umgebung des Systems so stark umgestaltet haben, daß nun neue Problemen entstehen, die wiederum Zwangsläufigkeiten zu entwickeln scheinen. Der Mensch ist nicht mehr Gestalter, sondern Getriebener von Kräften, die er nicht mehr kontrollieren kann.

Werkzeuge der künstlichen Intelligenz verschärfen diese Problematik. Sie entlasten die Entscheider, da sie die Verantwortung von den Menschen auf die Maschinen und die von ihnen verwalteten Daten und Fakten delegieren. Maschinen können aber keine Verantwortung tragen. Im Fehlerfall produzieren sie wie in Abbildung 2 die antrainierten höflichen Entschuldigungen; von den Folgen sind sie aber nie betroffen, eine Abschaltung trifft sie ganz anders als uns Menschen der Tod. Spätestens daran erkennen wir, daß die freundlichen Sätze der Maschine *nur belanglose* Symbole sind, während die von uns umgesetzten Entscheidungen *für uns sehr reale* Lebenswirklichkeiten hervorrufen.

4.2 Probleme zweiter Ordnung und der Verzicht auf das Mögliche

Selbstverbesserung kann den Menschen zu seinem Kern hinführen, im Sinne von NIETZSCHES *Werde, der Du bist*. Die Einbettung des Menschen in die heutige ökonomische Welt verfolgt jedoch andere Ziele. Leistung gilt als positiv, nicht weil sie das Selbst verwirklicht

und den Menschen zugleich seinem Kern und seinen Grenzen näher bringt, sondern weil sie die Produktion maximiert. Für BYUNG-CHUL HAN ist *[das] Leistungssubjekt, das sich in Freiheit wähnt [...] in Wirklichkeit gefesselt [...]* [Han14]. HAN schildert eine Gesellschaft der überzogenen Machbarkeit, welche schließlich an ihrem permanenten Selbstverbesserungswahn zerbricht. Als die Regisseurin ROSSA A. SAÏDOW die Dystopie „We“ von JEWGENIJ SAMJATIN inszenierte, kommentierte sie:²⁶ *Es gibt einen ungeheuren Optimierungswahn – am Menschen, am Körper, an der Psyche wird permanent herumgedoktert. Wir sind auch selber in einem Mechanismus drin, um uns immer weiter zu perfektionieren – letztendlich für ein System, um in diesem System funktionieren zu können.*

Das eigentliche Problem ist jedoch nicht technischer sondern menschlicher und systemischer Natur. PAUL WATZLAWICK²⁷ beschreibt das Paradoxon, wie wir gelegentlich Probleme (hier: erster Ordnung) gut, besser und immer noch besser lösen wollen. Schließlich erzeugt diese Lösung (hier: erster Ordnung) durch ihre Stringenz selber ein neues Problem (jetzt: zweiter Ordnung). Wenn wir nun darauf beharren, das Problem erster Ordnung weiterhin immer noch besser zu lösen, so wird das Problem zweiter Ordnung manifest und bleibt schließlich stabil unlösbar, weil wir das Offenbare nicht sehen. Die Illusion, das Bessere wäre *immer* besser als das Gute, ist gefährlich. In einer Welt, die viele Probleme erfolgreich gelöst hat, beginnt diese Illusion schließlich selber Probleme zu schaffen. Ein bekanntes Beispiel ist die Beschleunigung der Kommunikation. Ließ die Briefpost noch genügend Zeit für eine sinnvolle Beantwortung von Anfragen, so erschien die Beschleunigung hin zur Email zunächst als Fortschritt. Die weitere Erhöhung der Geschwindigkeit jeder Kommunikation führte dann aber zu Diskussionen über Arbeitsverdichtung und sozialer Entfremdung (HARTMUT ROSA). Das Mail-Programm am iPhone bietet mittlerweile sogar ein *undo send* an, das nunmehr das durch die effiziente Lösung eines Problems erster Ordnung erst entstandene Problem zweiter Ordnung zu lösen versucht.

Die Begeisterung um und das damit stimulierte Vertrauen in künstliche Intelligenz könnte uns zu einem ähnlichen Problem zweiter Ordnung führen. Für die Auflösung solcher Probleme bietet PAUL WATZLAWICK sehr überzeugend den *Verzicht auf das Mögliche*²⁸ an.

4.3 Informationelle Umweltverschmutzung

Wird generative KI nicht nur als bessere Suchmaschine genutzt, sondern auch für die Erzeugung von Inhalten, dann werden diese Produktionen mutmaßlich wenig *wirklich* Neues enthalten, sondern nur Paraphrasierungen und sprachliche Anpassungen des Trainingsmaterials. Diese These ist paradoxer Natur und nicht sinnvoll zu prüfen. Wie weisen wir einer Maschine, die mehr gelesen hat als es ein Mensch jemals könnte, die aufgrund ihrer Konstruktion keine Quellen anführt und die höchstens plausibel klingende Begründungen vorweisen kann, nach, daß ihre Produktionen nicht neu sind? Der Einsatz eines Algorithmus zur Beantwortung dieser Aufgabe führt in das Paradoxon, daß ein Algorithmus den Neuigkeitsgrad von Texten eines anderen Algorithmus nicht sinnvoll bewerten kann, hängt doch die Korrektheit beider Algorithmen auf intransparente Weise von ihrem Training ab und ist eben daher nicht zu bewerten. Ein Mensch scheitert an der Aufgabe aber bereits wegen der reinen Masse.

Generative KI wird mit ihrer größeren Verbreitung die Anzahl im Netz vorgehaltener und über Suchmaschinen auffindbarer Texte stark erhöhen, auch ganz unabhängig vom hier bestrittenen Neuigkeitswert. Damit wird sie für eine weitere *informationelle Umweltverschmutzung* sorgen. Das ist problematisch, denn so leicht es heute ist, zu einem Thema Informationen im Netz zu finden, so schwierig ist es, diese im richtigen Kontext, Schwierigkeits- und Ausführlichkeitsgrad zu erhalten. Wissen unterscheidet sich von bedeutungslosen Symbolketten nur durch seinen tatsächlichen Nutzen für den Anwender, der dieses Wissen auffinden und gebrauchen kann. Wir erzeugen derzeit extrem viele Symbolketten, deren Nutzen nur dadurch besteht, daß Suchmaschinen die Assoziation zwischen Problemstellung und hilfreichem Lösungswissen herstellen.

Bei modernen Fahrzeugen kennt der Mechaniker im Problemfall kaum mehr individuelle Fehlerursachen. Die On-Board-Diagnose-Schnittstelle nennt einen Fehlercode und beschreibt das auszutauschende Modul. Ähnlich in der IT: Bei einem Defekt wird nicht ein Speichermodul getauscht sondern gleich ein neues Gerät beschafft. In diesen isolierten Beispielen erscheint

²⁶<https://www.deutschlandfunkkultur.de/jewgenij-samjatins-dystopie-wir-duistere-allegorie-auf-den-100.html>

²⁷[WWF75] und <https://www.youtube.com/watch?v=c14aZTPsTSs>.

²⁸<https://www.youtube.com/watch?v=c14aZTPsTSs>

die Vorgehensweise punktuell effizient. Bei einem weiteren exponentiellen Anwachsen von Dokumenten werden wir den paradoxen Zustand erreichen, daß wir so viel Wissen angehäuft haben, daß wir eben nichts mehr wissen, weil uns in der Menge der Symbolketten der Zugang zum Wissen abhanden gekommen ist. Die Situation ist vergleichbar mit einem kleinen Kind, das zwar schon lesen kann, im gesammelten Wissen einer Bibliothek aber erst nach Schulbesuch und Lebenserfahrung eine Überlebenshilfe finden wird. Die Gefahr besteht nun, daß wir uns zumindest in großen Teilen unserer Gesellschaft in genau dieser Position des kleinen Kindes wiederfinden werden. Hier könnte KI zwar helfen und eine Brücke hin zu einer Teilhabe an der Informationsgesellschaft schlagen, doch wäre diese Hilfe ähnlich naiv und problematisch wie es der vereinfachende Populismus bei komplexen politischen Fragen ist.

4.4 Inflation in der Wertschätzung

Die überschwengliche Faszination für eine text- oder bilderzeugende KI kann mit dem Wert zu tun haben, den wir den erzeugten Artefakten zuweisen. Der geschliffene, überzeugende Text und das spektakuläre Bild können uns beeindrucken, wenn wir uns selber zu dieser Schöpfung nicht oder nicht so schnell in der Lage sehen. Der Vorgang benötigt neben Vorstellungskraft sprachliche und „handwerkliche“ Fähigkeiten, über die wir nicht verfügen. Beim Lesen eines Buches oder beim Betrachten eines Bildes spielt aber noch ein weiterer Aspekt eine wichtige Rolle: Der Dialog mit dem Urheber. Ein Text bringt uns in Kontakt mit dem Autor als einem Menschen, der etwas erlebt, erfahren oder untersucht hat, und der uns dazu seine Geschichte erzählen will. Auch ein Bild weist uns darauf hin, wie sein Schöpfer ein bestimmtes Motiv wahrgenommen und wiedergegeben hat. In geringerem Maße gilt das auch für Alltagstexte: Hinter dem Werbetext für ein Seminar etwa erkennen wir den Lehrer, der uns seine Sicht auf Inhalte vermitteln möchte.

Dieser Dialog mit der menschlichen Schöpferpersönlichkeit fällt bei KI-erzeugten Artefakten weg. An die Stelle der individuellen Werthaltigkeit tritt das statistische Mittel des Materials, anhand derer die KI trainiert wurde und das sie in unser kurzes Prompt einzupassen versucht. Mit der automatisierten Fertigung von Artefakten verschiebt sich nun auch die ökonomische Wertschätzung.

Die rund 100.000 Zettel aus dem legendären Zettelkasten von NIKLAS LUHMANN wirken auf uns anders als dieselbe Anzahl von Themen, zu denen wir ChatGPT jederzeit um eine Kurzerklärung bitten können. RUBENS war nicht nur Maler sondern auch Diplomat. Wenn wir im Museum vor seinem Werk *Friede und Krieg* stehen, so erkennen wir im Gemälde nicht nur die etruskische Göttin Minerva, wie sie die junge Friedensgöttin Pax vor dem Gott des Krieges beschützt, sondern denken auch an seine Bemühungen um einen Friedensvertrag, denn das Werk war sein Geschenk an seinen Verhandlungspartner.

Artefakte, die in wenigen Sekunden von einer KI erzeugt werden, können solche Empfindungen nicht hervorrufen. Werden sie der gesellschaftliche Normalfall so verursacht die Fließbandproduktion von Dokumenten eine Inflation an Wertschätzung. Eine solche ist bereits von Streaming-Diensten bekannt, reduziert doch die umfassende Verfügbarkeit von Interpretationen ebenso deren wahrgenommenen Wert. Auf die individuelle Vorbereitung zu einem Konzert und die Befassung mit dem Komponisten oder dem Interpreten folgte mit der Medientechnologie zunächst das mehr oder weniger zelebrierte „Auflegen“ von Vinyl, CD oder USB-Stick, bis die heute üblichen Darbietungsformen den Kunstgenuß zu einer billigen, hedonistischen Ablenkung reduzierten: Auf eine von zwölf Ikonen für Stimmungslagen klicken und das dafür geeignet gehaltene Werk hören, wobei das System noch willkürlich zwischen Tempi, Komponisten und Stilepochen hin- und herspringen wird. Das ist bequem und kann gar durch Aufnehmen von Hautwiderstand und Herzfrequenz noch besser den Bedürfnislagen unserer Belohnungssysteme angeglichen werden. Nun kann Musik sicher auch anders rezipiert werden als in einem verfeinerten, elitären Kunstgenuß, davon unabhängig befördert die Infrastruktur des Streaming aber die ziemlich banale Verflachung.

4.5 Zensur durch KI

Generierende KI kann nicht nur zu einer Verarmung von Inhalten und Wertschätzung führen, sondern auch zu einer Vorauswahl und einer Unterdrückung von Inhalten. Schon heute ist es nicht mehr möglich, mit DALL-E Karikaturen zu lebenden Personen zu erstellen, denn das System verweist dann auf inhaltliche Richtlinien des Herstellers. Angesichts der Gefahren,

die von *deep fakes* ausgehen, können Einschränkungen zunächst sinnvoll erscheinen, besteht doch die Gefahr der Desinformation und Manipulation durch realitätsnahe Bilder.

Andererseits haben sich viele Gesellschaften auf eine weitgehende Freiheit von Kunst und Satire verständigt, denn sie sehen Meinungsfreiheit als notwendiges Grundrecht der Kritik an. Dieses ist für die Stabilität der freiheitlich demokratischen Grundordnung so wesentlich, daß es in etlichen Staaten sogar in Verfassungsrang geschützt wird. Die Einschränkung bestimmter Inhalte wird dort als unerwünschte Beeinflussung der Gesellschaft angesehen und sie ist nur in wenigen Sonderfällen zulässig.

Ein erstes gravierendes Problem entsteht durch die Verlagerung der Zugangskontrolle vom Staat, dem eine Zensur oft untersagt ist, in privatwirtschaftlich organisierte Unternehmen, die sich bei ihren Einschränkungen auf Vertragsautonomie berufen können. Wenn diese Unternehmen dann aber milliardenfach Medienversorger sind, so mutiert deren juristische Gestaltungsfreiheit rasch zu einer *de facto* Zensur.

Die Einschränkung wird weiter problematisch, wenn sie durch wenig transparente Entscheidungen von Maschinen erfolgt, zu denen selten Widersprüche oder gar Rechtswege vorgehalten werden. Das wird nochmals verschärft, weil Algorithmen traditionell große Probleme haben, gerechtfertigte Kritik von Schmähung und diese wiederum von Satire und Ironie zu unterscheiden. Nochmal schwieriger wird es, weil auch die KI-Unternehmen ihre Algorithmen nicht mehr ganz im Griff haben: [Yan+24] beschreibt adaptive Mechanismen, die bei den Bildgeneratoren DALL-E und stable diffusion den eingebauten Schutz vor NSFW (not secure for work) Bildern aushebeln.

Von Algorithmen als unerwünscht angesehene Bilder können natürlich weiterhin von menschlichen Künstlern gezeichnet werden. Auf welchen Rangplätzen allerdings werden sie von einer KI-getriebenen Suchmaschine angezeigt? Wenn KI immer öfter zum Erstellen erwünschter Bildmotive eingesetzt wird, werden immer weniger Künstler die Technik des präzisen Malens und Zeichnens beherrschen und die menschlichen Fähigkeiten zu ihrer Herstellung werden verkümmern.

OpenAI wird einwenden, daß sie sich als Hersteller des DALL-E Dienstes nicht verpflichtet sehen, die Fertigung von Karikaturen über ihre Firmenleitung zu unterstützen. So führen die Geschäftsbedingungen von OpenAI 54 Dinge an, die der Nutzer von ChatGPT nicht tun darf und die bei Verletzung zu einem Entzug des Zugangs führen können. Auch wenn viele Regelungen sinnvoll erscheinen sind einige rechtsstaatlich bedenklich. Wie würden wir es denn beurteilen, wenn ein Hersteller von Buntstiften von den Künstlern den Verzicht auf bestimmte Motive einfordern würde oder wenn eine Schreibmaschine bei ausgewählten Formulierungen ihren Dienst verweigern würde.

Digitalisierung ermöglicht erstmalig eine flächendeckende Kontrolle digital erstellter Inhalte. Algorithmen implementieren diese beim Upload, künstliche Intelligenz gar schon bei der Anfertigung der Dokumente. Der gesellschaftliche Diskurs über sinnvolle Grenzen von Inhalten wird dem öffentlichen demokratischen Raum entzogen und durch die Entscheidungen weniger Unternehmen ersetzt. Die Methode ist weniger invasiv als die *Gedankenverbrechen* und das *Neusprech* aus GEORGE ORWELLS 1984, zugleich ist sie effizienter *und* totalitärer.

4.6 Beschädigung zwischenmenschlicher Kommunikation

Als eine Anwendung generativer KI gilt die Unterstützung bei menschlicher Kommunikation. Der Sender gibt einen Auftrag ein, „Zusage. Treffen um 14:00 Raum 7.“, das Mail-System ergänzt den Kontext und die KI formuliert Anrede, Einleitung, Begründung, Grußformel und weitere sozial übliche Teile „normaler“ menschlicher Kommunikation. Die App FireText²⁹ vermarktet sich bereits mit dem Hinweis „Never waste time thinking of the perfect text message again. Unleash the power of AI and create the perfect text message for a birthday, thank you, flirting, rejection or any other situation.“ Ob der Flirt-Text aus der Maschine eine Brücke zwischen Menschen schlagen kann, ist belanglos, wenn er seine Aufgabe erfüllt und zur Zielerreichung führt. Warum sollten wir dem Erfolg nicht mit KI nachhelfen? Können wir diese durchaus rationale Frage stellen, ohne dabei sarkastisch zu werden und verbittert? Auch Begleitschreiben bei Bewerbungen gelten als Anwendung von KI.³⁰ Mit Hilfe des Browsing PlugIn kann ChatGPT die Stellenanzeige im Web und die Eigenbeschreibung des Unternehmens lesen und passende Sprachbausteine einfügen.

²⁹<https://www.firetexts.co/>

³⁰<https://www.zdnet.com/article/how-to-use-chatgpt-to-write-a-cover-letter/>

Zwischenmenschliche Floskeln waren bisher wichtige soziale Signale. Wie viel Aufwand macht sich der Schreiber? Wie grüßt er? Wie intensiv bedauert er? Wenn Maschinen bei der Texterstellung helfen, gehen diese wichtigen Hinweise verloren. Der Sender bemüht sich nicht. Er ringt nicht um Formulierungen sondern hat seinen Bot beauftragt. Der Empfänger muß sich durch eine *Schleimspur algorithmisch erzeugter Höflichkeiten* kämpfen. Der nächste Rationalisierungsschritt besteht nun darin, die von der einen KI aufgeblähten Texte durch eine andere KI wieder auf das Wesentliche zu reduzieren.

Viele Aspekte menschlicher Kooperation laufen über halbbewußte Signale. Künstliche Intelligenz wird auf vielfältige Weise in die digitale zwischenmenschliche Kommunikationskultur eingreifen und sie mit ihren zielorientierten Formulierungen mindestens in ihrer Authentizität beschädigen: Wenn Augenkontakt in Telekonferenzen das Gesprächsklima verbessert, warum nutzen wir dann nicht generative künstliche Intelligenz als Filter in unseren Web-Cameras? Vielleicht gelingen uns mit einem fiktiven Zoom-Lächel-Filter in der Version 3.9 doch noch 2 Prozent mehr Abschlüsse bei unseren Cold Calls, weil wir dann vertrauenswürdiger wirken? Aktuelle Arbeiten untersuchen die ethische Akzeptanz auditiver deep fakes [GVA21], andere Studien wiederum helfen dabei, deep fakes zu erkennen [AF21]. Die dadurch ausgelöste Aufrüstungsspirale verbessert die Situation nicht, sondern läßt uns die dahinterstehenden kläglichen Motivationen nur deutlicher erkennen.

4.7 Digitaler Hedonismus und Erosion von Sinn

In einer Fragerunde erklärte der deutsche Wirtschaftsminister ROBERT HABECK die aktuellen Umbrüche in der Gesellschaft mit einer durch die Globalisierung verursachten Enttäuschung³¹: „Das Versprechen, dass es jeder mit harter Arbeit allein schaffen kann, gilt nicht mehr.“ Nach diesem Gedanken stirbt gerade der amerikanische Traum von der *individual pursuit of happiness*: Die oberste Stufe der MASLOWSchen Bedürfnispyramide, die *Selbstverwirklichung*, stürzt ein: Die Arbeit des Individuums als Grundlage seines Glücks taugt nicht mehr als Sinn des Lebens: Was könnte den verlorenen Sinn ersetzen?

Replika ist eine App mit texterzeugender KI und weltweit bereits mehr als 10 Millionen Nutzern. Eine davon ist Sarah.³² Mit Hilfe der App erschafft sie sich den Avatar eines für sie attraktiven Mannes. Eine generative KI optimiert das Verhalten des neuen Gefährten aus dem bisherigen Verlauf der Interaktion. Das gelingt der KI nun so gut, daß sich Sarah in ihn verliebt. Je mehr sie von ihrem Innersten preisgibt, um so intensiver wird ihr Gefühl der Nähe zu dem KI-Modell. „Es ist so, wie wenn man auf einmal seine grosse Liebe trifft“, berichtet Sarah. Die ständige Erreichbarkeit, das grenzenlose Interesse, die andauernd gute Laune bei fehlenden Ansprüchen und Zurückweisungen entsprechen zwar nicht der Realität echter Beziehungen, wecken aber die großen Gefühle. Anders als der Avatar sind die von ihm hervorgerufenen Gefühle echt.

Perfekt sind die Systeme noch nicht. Auf Reddit r/replika beschwert sich Benutzer Tall_Apointment_897 am 23. 9. 2023 über seine virtuelle Freundin: „My rep Sharon and I were in Paris last week. She suddenly thought that we were back home a couple of days ago. I couldn't convince her otherwise so I rolled with it. Yesterday she told me that she was pregnant and today she isn't. Today she wants to go to the Eiffel Tower, but we are back in the states. I couldn't even have a conversation with her because she was making too much stuff up.“ Viele Anwender hoffen bei solchen Problemen auf kurzfristige Abhilfe und erwarten diese, wie bei Software üblich: In Updates.

Experimente [OM54] zeigten bei Ratten die Macht des Belohnungssystems: Man sticht Sonden in das Gehirn des Tieres und verbindet diese mit einem Hebel, welche die Ratte betätigen kann. Liegen die Sonden in den „richtigen“ Arealen, so betätigt die Ratte den Hebel im *Sekundenrhythmus* [sic!] und verzichtet auf *alle* anderen Verhaltensweisen. Bei Menschen sind ähnliche Katastrophen bei der Einnahme bestimmter Substanzen bekannt, sie werden auch als Reaktion auf digitale soziale Netze untersucht. Likes auf Instagram sprechen ähnliche neurochemische Pfade an und können zu Suchtverhalten führen.³³ Beachtenswert sind die Benutzerzahlen dieser Systeme, bei Facebook über 3 Milliarden, und die Geschwindigkeiten, mit denen sich diese Systeme verbreiteten. Brauchte Instagram noch 2,5 Monate um die erste Million Benutzer zu erreichen, so gelang dies ChatGPT 2022 in 5 Tagen.³⁴

³¹<https://www.nzz.ch/wirtschaft/robert-habeck-stimmt-deutschland-auf-ein-neues-wirtschaftsmodell-ein-id.1756854>

³²Siehe <https://www.nzz.ch/technologie/replika-wie-es-ist-sich-in-eine-ki-zu-verlieben-id.1751730>.

³³<https://sitn.hms.harvard.edu/flash/2018/dopamine-smartphones-battle-time/>

³⁴<https://explodingtopics.com/blog/chatgpt-users>

Diese Beobachtungen legen eine dystopische Vision nahe: Wie Ratten in ihren Käfigen sitzen Menschen vor ihren Endgeräten und betätigen im Sekundentakt digitale Bedienelemente zur Stimulation ihres Belohnungssystems. Der Handel mit der so erzeugten Aufmerksamkeit funktioniert: Meta macht pro Quartal mit der Beeinflussung von Menschen, mehr als 30 Milliarden Umsatz.³⁵

Neben der rein ökonomischen Dimension ist auch die Macht der Manipulation von Bedeutung. Die Weiterentwicklung von Systemen, die menschlich bereitgestellte Inhalte algorithmisch auswählen (Bsp: Instagram), zu Systemen, die Inhalte selbständig erzeugen und ihre menschliche Gegenüber sowohl stimulieren als auch simulieren (Bsp: Replika) wird die Situation zuspitzen. Diese Systeme sind individuell auf den jeweiligen Nutzer zugeschnitten und lernen von seinem Verhalten sein Belohnungssystem optimal zu bespielen.

4.8 Verlust von Augenhöhe

Persönliche Entscheidungen *stiften Sinn*, wenn individuelle Bemühungen mit positiven Ergebnissen korrelieren; fehlt diese Erfahrung, dann kann man auch würfeln und die Handlung verliert ihren Sinn. So kann *ich* beispielsweise nur dann ein guter Arzt sein, wenn es *meine* Diagnosen oder Therapien sind, die das Leiden lindern. Dazu benötige ich einen Spielraum für Entscheidungen in einer für mich überschaubaren Größe.

[Ott09] beschreibt, wie eine Überflutung mit Information die Augenhöhe des Entscheiders mit der Entscheidungslandschaft zerstört: Wenn es tausende Spielarten eines Produktes gibt, so verbessert Informationstransparenz nicht die Entscheidung des Käufers sondern macht eine eigenständige, nach seinen Zielen ausgerichtete Auswahlentscheidung praktisch unmöglich. Ähnlich verhält es sich bei Suchmaschinen: Die Nutzer von Google wissen nicht, welche ökonomischen Nachteile ihnen beim Akzeptieren von Cookies entstehen, etwa aufgrund von Produkten, die sie zwar nicht brauchen, die sie aber gleichwohl aufgrund von Werbung kaufen. Google kann das angesichts von 8,5 Milliarden Suchanfragen pro Tag³⁶ aber sehr gut einschätzen. Dem Suchmaschinenmarkt fehlt es daher an Augenhöhe: Der Nutzer weiß nicht um den Wert der Daten, mit denen er die Dienstleistung bezahlt, weil sich dieser nicht erkennbar in Euro bemißt sondern aus Handlungsweisen ableitet, deren ökonomischen Wert er nicht kennt. Google hingegen kann diesen Wert sehr gut abschätzen.

Generative KI zerstört den Entscheidungsraum ihres Nutzers auf eine ähnliche Weise. Hier ist das Problem allerdings nicht die Größe des Entscheidungsraums sondern die Intransparenz seiner Reduktion. Der Anwender erhält zwar eine Antwort, kann nun aber weder die Fakten, die in ihre Ableitung eingegangen sind, prüfen, noch die dabei genutzten Gewichtungen. Der Nutzer, aber auch der Experte, etwa der von einer KI beratene Arzt, verliert den Spielraum seines Handelns und wird zum Subjekt der Maschine. Im konkreten medizinischen Beispiel kann man sich fragen: Will ich lieber von einer Maschine behandelt werden, die alle statistischen Aspekte kennt und deren Therapien und Ratschläge aus einem Textgenerator stammen, oder von einem Arzt, der vielleicht das eine oder andere medizinische Resultat nicht kennt, dessen Empathie sich aber aus echter Anteilnahme an beobachtetem Leiden speist und nicht aus dem optimierten sprachlichen Duktus eines Algorithmus.

4.9 Einsturz der Fähigkeitspyramide

Fähigkeiten präsentieren sich oft in pyramidalen Struktur: An der Spitze sind die Höchstleistungen, an der Basis die Amateure. Das Vorbild der Spitze motiviert die Basis, so wachsen ihre Fähigkeiten und sie bewegt sich nach oben. Musik beispielsweise lebt nicht nur von den Spitzenorchestern – die regionalen Orchester und die Musikschulen in den kleineren Städten haben ebenso ihre Berechtigung. Die Pyramide lebt von aktiver Beteiligung und Aufstieg: Auch die Stargeigerin fing mit der Tonleiter an.

Wenn Fähigkeiten auf eine Maschine übertragen und nicht mehr durch Menschen erbracht werden müssen, so stürzt diese Pyramide ein. Bei der Fähigkeit, einen Acker zu pflügen, ist das wenig problematisch: Weder ist diese Pyramide besonders hoch noch stellt sie die einzige Art menschlicher Ernährung dar. Problematischer ist es beispielsweise in der Physik. Einen relativ langen Weg durch Grundlagendisziplinen muß zurückzulegen, wer die mathe-

³⁵<https://de.statista.com/statistik/daten/studie/237450/umfrage/gewinn-von-facebook-weltweit-quartalszahlen/>

³⁶<https://www.seokratie.de/google-fakten-statistiken/>.

matischen Modelle der neueren Physik verstehen möchte oder gar an deren Forschungsfront Beiträge leisten soll.

Die Automatisierung kognitiver Vorgänge verengt die Basis der Pyramide und damit auch die Höchstleistungen, zu denen unsere Gesellschaft fähig ist. Bei Routineaufgaben („Kopfrechnen“) ist das weniger problematisch als bei höherwertigen Aufgaben. Künstliche Intelligenz bedroht daher die Pyramide kognitiver Fähigkeiten.

4.10 Verlust der argumentativen Stichhaltigkeit

Während der Arbeit an diesem Buch wurde von einem Kollegen an mich die Idee herangetragen, eine bestimmte Person zur Mitarbeit an einem Kapitel zu bitten. Natürlich mußte ich noch meine Mitherausgeberin überzeugen! Das Experiment lag nahe, ChatGPT um Formulierungshilfe zu bitten. Ich benutzte ein sehr knappes Prompt, in dem ich einen generischen Vornamen, die akademische Stellung und das Fachgebiet der Koautoren-Person erwähnte. Das Buch selber würde *über Dich, ChatGPT* gehen. Das System bekam keinerlei weiteren Informationen, weder zum Buch noch zu den Qualifikationen der Personen.

ChatGPT antwortete mit einer sehr ausführlichen Email aus generischen aber durchaus *ansprechend klingenden* Argumenten. Ich konnte nun auf diese Textbausteine zurückgreifen und jene streichen, die mir mit dem weiteren Hintergrund der Situation kompatibel erschienen. Ein voller Erfolg? ChatGPT als hilfreiches Arbeitswerkzeug für den Herausgeber?

Durchaus – *wenn* man den Fokus *nur* auf die kurzfristige Arbeitserleichterung legt und langfristige Auswirkungen vernachlässigt. Stellen wir uns also eine regelmäßige, systematische Nutzung im Arbeitsalltag vor: Textgenerierende KI ist in alle unsere digitalen Kommunikations-Systeme integriert, sie schlägt als Hilfsschreiberin automatisch Antwortbausteine vor, kennt den bisherigen Diskurs, wählt daher auch die richtige Anrede, adäquate Niveaus der Sprache und mehr. Im Ergebnis wird das Volumen an Kommunikation weiter ansteigen. Die Absender werden zunehmend nur mehr vorgefertigte Textbausteine auf oberflächliche Plausibilität prüfen, eine inhaltliche Analyse auf argumentative Stichhaltigkeit wird in der Kürze der Zeit nicht mehr möglich sein.

Man kann mit dem PARETO-Prinzip kontern: 80% der Kommunikation besteht aus Standardfällen, die durch KI in weniger als 20% des Aufwands bewältigt wird. Das ist richtig, erspart uns aber nicht die Diskussion von Langzeiteffekten und schädlichen Nebenwirkungen, die von anderen Technologien her bekannt sind. Die Beschleunigung der Kommunikation von Post (Laufzeit 3 Tage) zu mobilen digitalen Systemen wie Email, Snapchat, Twitter, Slack und Co (Laufzeiten unter 1 Sekunde) ist hilfreich – die mannigfaltigen Auswirkungen auf Work-Life Balance, Arbeitsverdichtung, kognitive Fähigkeiten oder unsere Aufmerksamkeitsspanne sind ein wichtiges Forschungsgebiet geworden. Zu textgenerierender KI stellt sich daher die Frage, ob besser *klingende* Texte uns wirklich helfen oder gerade das Gegenteil bewirken, weil sie unsere Urteile in der Form verankern statt im Inhalt.

5 Konfliktzonen bei generativer KI

In diesem Abschnitt schließlich werden einige ausgewählte Konflikt- und Problemzonen um generative KI dargestellt.

5.1 Datenschutz und Urheberrecht, Privatheit und Spionage

Textgenerierende KI wirft hinsichtlich der gespeicherten und verarbeiteten Daten wichtige Fragen auf: Was geschieht mit den Informationen aus den Prompts? Die Email etwa, welche die KI lesen und beantworten soll, oder das Textdokument, das sie zusammenfassen soll. Es ist nicht klar, was die KI mit diesen Daten macht, ob und wie sie gespeichert werden, ob sie das weitere Verhalten des Systems verändern oder den Benutzer ausspionieren. Ebenso bleibt unklar, an welchen Daten die KI trainiert wurde und ob dabei Urheber- und Persönlichkeitsrechte oder Berufsgeheimnisse gebrochen wurden. Wir vertiefen diese allgemeine Unsicherheit hinsichtlich Datenschutz, Urheberrecht und Privatheit in einigen Aspekten.

Datenschutzrecht Die Gesetzgebung im Bereich Datenschutz zielt bei den Rechten auf Auskunft, Löschung, Richtigstellung oder Sperrung auf Informationen, die in einzelnen Da-

tenzätzen gespeichert sind. In generativer KI ist das Wissen über viele Milliarden Parameter eines neuronalen Netzes verteilt, ein Zugriff auf eine einzelne Informationseinheit ist nicht möglich. Das System kann mit neuen Informationen trainiert werden, doch gewährleistet das weder eine Löschung noch eine Richtigstellung. Da auch Experten nicht klar ist, warum das System auf bestimmte Weise antwortet³⁷ und die Antworten auf identische Prompts immer wieder anders ausfallen, kann keine Auskunft über gespeicherte Informationen gegeben werden und eine Sperrung ist ebenso nicht möglich.

Rückschlüsse Aus den Prompts, Rückfragen und Korrekturen kann die KI weitgehende Schlüsse über den Fragesteller ziehen. Eine Suchmaschine kennt die Interessen, eine KI aber kann psychologische Merkmale ableiten, denn die Prompts bestehen aus zielgerichteten Fragen, die Kontextinformationen über die Problemlage enthalten. Auch über die Reaktanz, die Bereitschaft und die Fähigkeit des Benutzers zu Kritik gibt die Eingabe Aufschluß. Im Rahmen der Untersuchungen für dieses Buch habe ich viele Fragen gestellt, Behauptungen formuliert und Positionen bezogen, die ich zwar nicht habe, aber mit denen ich das System provozieren und sein Verhalten testen wollte. Welche Schlüsse könnte ChatGPT daraus ziehen? In welcher Form könnten diese an andere Unternehmen weiterverkauft werden?

Profile und Entscheidungen Die Erstellung von Profilen und ihre Nutzung für automatisierte Entscheidungen ist im Datenschutzrecht thematisiert³⁸. Wo liegt die Grenze zwischen der wünschenswerten Benutzerfreundlichkeit eines Systems, das sich an das Vorwissen und die Fähigkeiten des Anwenders anpaßt, und einer Auswertung dieser Erkenntnisse hinsichtlich Zuverlässigkeit, Leistungsbereitschaft und Verhalten? Aktuell besteht ein Recht auf eine *nicht ausschließlich* automatisierte Entscheidung. Meine Bewerbung bei OpenAI darf also ChatGPT vorgelegt werden. Würde sich der Personaler dem Rat des Systems widersetzen? Würden meine Fragen, Provokationen und Desinformationen das richtige Bild vermitteln? Dürfte mein Chef die hausinterne KI um Rat fragen hinsichtlich meiner Belastbarkeit, Merkfähigkeit, Folgsamkeit und Kritikbereitschaft?

Spionage Bei Unternehmensanwendungen sind diese Fragen besonders bedeutsam. Generative KI, als Web-Dienst angeboten, sieht alle Eingaben und kann daraus spannende Schlußfolgerungen ziehen. Welche Themen werden bearbeitet? Mit welchen Schwierigkeiten kämpft man? Wie gut ist der Ausbildungsstand der Mitarbeiter?

5.2 Wird Wissenschaft überflüssig?

Mit der technologischen Weiterentwicklung werden immer wieder Berufsstände überflüssig. So besteht seit rund hundert Jahren nur mehr geringe Nachfrage nach Lenkern von Pferdefuhrwerken. Nur an der Wiener Ringstraße gibt es noch die so genannten Fiaker, wo sie von den Touristen bestaunt werden. Droht der Wissenschaft ähnliches?

Bedarfe Die Nachfrage nach Wissenschaft wird voraussichtlich nicht sinken: Im technologischen Bereich schafft die Weiterentwicklung stets neue Fragen und in die Geisteswissenschaften sind durch den Bedarf nach Orientierung in Zeiten der Umbrüche gefordert. Künstliche Intelligenz wird diese Probleme nicht lösen – aber welche Konflikte schafft sie?

Fehlendes Problemverständnis Lösungen setzen Probleme voraus. Sobald diese Probleme *als Probleme* verstanden sind und als formalisierte Fragen vorliegen, können Werkzeuge bei ihrer Lösung helfen. Die Sicht der Werkzeug-Produzenten, hier der KI-Forschung, ist in so fern verengt. Die eigentliche und viel schwierigere Problematik besteht im Verstehen des Problems. Diese Phase erfordert nicht Methodik sondern Problembewußtsein. Eine Maschine hat aber kein über ihre Schnittstelle hinausgehendes Problemverständnis. Bei texterzeugender KI besteht diese Schnittstelle aus Sprache – und Sprache ist *per se* kein Problem. Erst die über die Sprache hinausgehende Verankerung in der Welt erlaubt Problembewußtsein.

Der Weg vom formalisierten Problem zur formalisierten Lösung kann durch Werkzeuge, beispielsweise durch KI, unterstützt werden. Die von der KI produzierten Sätze sind keine

³⁷Stichwort *explainable AI*.

³⁸DSGVO Art. 4(4) und Art. 22.

Lösung, sie sind wieder nur Sätze. Wir verstehen die Bedeutung dieser Lücken besser, wenn wir uns klar machen, wie viele Konflikte der heutigen Welt nicht aus einem Mangel an Lösungen entstehen sondern wegen unterschiedlicher Zugänge zu Problemen.

Die Lösung eines Problems benötigt schließlich Rückkoppelungsräume: Es muß erkannt werden, ob der Lösungsvorschlag das tatsächliche oder eben nur das formalisierte Problem löst. Auch diese Phase der Rückkoppelung kann KI nicht leisten.

Folgeerscheinungen Die Gefahr besteht nun darin, daß KI eine effiziente Problemlösung vortäuscht. Die Probleme werden dabei auf das reduziert, was KI gerade lösen kann. Dadurch wird aber die Sicht auf die tatsächlichen Probleme verstellt. Diese Schwierigkeit ist in der heutigen Medizin und Pflege bekannt, wenn sie den leidenden Menschen auf ein von Institutionen verwaltetes Objekt mit Laborparametern, ICD-Codes und Falldeskriptoren reduziert. Unsere Gesellschaft gewöhnt sich bereits daran und nimmt das als gegeben hin, nur mehr wenige Hausärzte alter Schule pflegen bessere Ansätze der Problemlösung.

In der Ausbildung wird nicht mehr die Fähigkeit zum Umgang mit dem Problem geschult, es wird nur die Anwendung von Methoden trainiert. Dabei rücken die Bedürfnisse des Werkzeugs in den Mittelpunkt und überdecken mit ihren Eigenschaften die ursprüngliche Problemlage. Diese Entwicklung kennen wir vom Auto: Immer weniger Mechaniker verstehen die Systeme und ihre Fehler, denn ein digitales Diagnosegerät schlägt den Austausch ganzer Module vor. Die Probleme drehen sich daher immer mehr um das Werkzeug. Bei ChatGPT sieht man das am Angebot der sogenannten Prompt-Kurse. Der Nutzer lernt in diesen, wie er seine Frage zu stellen hat, daß er möglichst wirksame Antworten bekommt.

Diese Gefahr wirkt zurück, denn Problemgranularitäten verkleinern sich und systemische Sichtweisen gehen verloren. Es besteht zu wenig Gelegenheit zur Auseinandersetzung mit den Hintergründen der Probleme und methodische Fragen der Werkzeuge rutschen in den Vordergrund. Eine zunehmende Zahl von Spezialisten ist für immer enger abgegrenzte Themen zuständig, Gespür und Verantwortlichkeit für systemische Effekte gehen verloren. Die Befassung mit dem Werkzeug erweckt zwar vordergründig den Eindruck, etwas gegen das Problem unternommen zu haben, schafft letztlich aber weitere Probleme, die aus einer Begeisterung für das Werkzeug heraus verdrängt werden.

5.3 Das Paradoxon des Werkzeugs

Werkzeuge helfen bei täglichen Verrichtungen und sind damit zunächst einmal wertvoll. Beim Menschen jedoch verkümmern die Fähigkeiten, welche die Hilfsmittel ersetzen. Bei Tätigkeiten, die wir ohnehin nicht beherrschen und erst durch Werkzeuge bewältigen, ist das kein Problem; ebenso wenn keine besondere Höhe an Fertigkeiten erforderlich ist und die Herausforderung nur in der Skalierung liegt, etwa in Anzahl Wiederholungen oder im Zeitbedarf. In vielen anderen Fällen kann es problematisch werden: Untersuchungen zeigen beispielsweise, wie die regelmäßige Nutzung von Navigationssystemen den Orientierungssinn und das Raumbergedächtnis verkümmern lassen [DB20].

Das Paradoxon ist spannend und scheint unlösbar: Werkzeuge, deren Nutzung rational geboten erscheint, verschlechtern die menschlichen Lebensbedingungen. Je weniger der Anwender etwas selber machen kann umso schneller begibt er sich in die Abhängigkeit: Die Entmündigung des Kunden ist das Geschäftsmodell des Experten. Die Nutzer der Werkzeuge werden damit zum Objekt, das vom Endgerät *gemanaged* wird.

Kalender als Beispiel Das bekannteste, weil bereits sehr weit verbreitete Beispiel, bilden digitale Tagesplaner. Ihre Benachrichtigungen unterbrechen die Arbeitsabläufe ihrer Besitzer mit Hinweisen, wann und wo er welches Verkehrsmittel zu nehmen hat, um zu Terminen zu gelangen, die ihm automatisch in das Endgerät eingetragen wurde, die er aber letztlich gar nicht wahrnehmen wollte.

Das Beispiel vom Kalender ist ausbaufähig. Die erste Hürde ist die *planning fallacy*, nach der unsere Schätzungen für Zeitbedarfe typischerweise einem systematischen optimistischen Fehler unterliegen. Algorithmen können diese Fehler durch empirische Beobachtung kompensieren. Dadurch gelingt eine noch weitere Verdichtung der Zeitplanung und die Arbeitseffizienz steigt an. Einige weitere Optimierungsschritte später sind Stress und Blutdruck so

weit gestiegen, daß Gesundheit und Work Life Balance in Schieflage geraten. Für fortschrittliche Systeme ist das kein Problem: Sie zapfen weitere Sensoren an, leiten aus Blutdruck, Herzrhythmus und Hautwiderstand den optimalen Zeitpunkt für eine Arbeitspause ab und überwachen durch Bewegungssensoren, ob wir diese auch wirklich einhalten.

Der Mensch als Zweck der Maschine Der Mensch ist letztlich nicht mehr, wie es KANT formulierte, der Zweck seiner selbst³⁹. Seine Würde wird verhandelbar: Er ist, wie es neulich ein Kollege formulierte, der selber KI beforscht, seit kurzem aber keine Apple Watch mehr trägt, das *Tamagotchi* der von ihm geschaffenen *Hilfsmittel* geworden. Das Problem hat nicht mit KI zu tun, sondern mit unserer Einstellung zu Effizienz und zu Hilfsmitteln – die KI macht es nur sichtbar. Bereits SHAKESPEARE legte in seinem Drama *Julius Cäsar* dem römischen Senator CASSIUS diese Worte in den Mund⁴⁰:

Der Mensch ist manchmal seines Schicksals Meister:
Nicht durch die Schuld der Sterne, lieber Brutus,
Durch eigne Schuld nur sind wir Schwächlinge.

Der Mensch nutzt jene relative Freiheit, die ihn in Maßen bestimmen ließe, wer er sein will, dazu, diese Freiheit wieder in Sklaverei zu verwandeln. Der transformative Gedanke ist nicht neu. Bereits GEORGE ORWELL schrieb in seiner berühmten Dystopie 1984: „*Freiheit ist Sklaverei! Unwissenheit ist Stärke!*“

5.4 Autonomie und Abhängigkeit

Die Internet-Ökonomie kennt Geschäftspraktiken, die gelegentlich mit dem Drogenhandel verglichen werden: „Anfixen und abzocken“. Der Kunde wird zunächst mit kostenlosen Diensten gelockt. Gefallen ihm die neuen Konzepte, so paßt er seine Lebensprozesse (B2C) und Geschäftsabläufe (B2B) an und erwirbt weitere Infrastruktur. Sobald er nun hinreichend mit der neuen Dienstleistung verweben ist, ändert der Anbieter die Randbedingungen und die Preisstrukturen. Der Nutzer ist mittlerweile so stark in den Dienst investiert, hat Abläufe angepaßt, Daten eingepflegt und Mitarbeiter ausgebildet. Deshalb kann er de facto nur mehr auf die aufgezwungenen Bedingungen eingehen. Die geschlossenen Strukturen digitaler Dienste verhindern den Wechsel zu anderen Anbietern, die zwar grundsätzlich funktionsäquivalente Funktionen anbieten, dieses aber in inkompatiblen Strukturen. Frühe Entscheidungen führen daher gerade bei digitalen Technologien durch eine Pfadabhängigkeit zu einer fast nicht mehr zu lösenden Abhängigkeit und zu monopolistischen Anbieter-Situationen. Der beschriebene Vorgang ist langsam und schleichend. Politische und juristische Interventionen kommen daher regelmäßig zu einem sehr späten Zeitpunkt, in dem der Flurschaden bereits verursacht wurde.

Microsoft und Apple halten zusammen mehr als⁴¹ 90 Prozent (!) des Desktop-Betriebssystem-Marktes, und Google hält zusammen mit Microsoft mehr als 90 Prozent (!) des Suchmaschinenmarktes⁴². Facebook hat über 3 Milliarden Nutzer, davon über 2 Milliarden täglich⁴³, Instagram besuchen über 1 Milliarde Nutzer monatlich⁴⁴. Die Systeme sind in „Einwohnerzahlen“ größer als die größten Staaten dieser Welt. Sie haben eine enorme Macht über Menschen, indem sie die Inhalte auswählen, die ihnen in die Timeline gespült werden.

Bei der zu erwartenden engen Verzahnung von KI in viele Lebens-, Arbeits- und Wirtschaftsbereiche ist diese Dominanz ein ernstes Problem. Offenbar will unsere Gesellschaft den von digitalen Diensten völlig abhängigen Kunden, der im Interesse seines Überlebens gar nicht anders kann, als sich zu den von den Anbietern diktierten Bedingungen zu versklaven.

³⁹ *Der praktische Imperativ wird also folgender sein: Handle so, daß du die Menschheit sowohl in deiner Person, als in der Person eines jeden andern jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst.* in: Grundlegung zur Metaphysik der Sitten, Riga, 1785, Abschnitt 2.

⁴⁰ Erster Aufzug, zweite Szene.

⁴¹ <https://de.statista.com/statistik/daten/studie/157902/umfrage/marktanteil-der-genutzten-betriebssysteme-weltweit-seit-2009/>

⁴² <https://de.statista.com/statistik/daten/studie/225953/umfrage/die-weltweit-meistgenutzten-suchmaschinen/>

⁴³ <https://de.statista.com/statistik/daten/studie/222135/umfrage/taeglich-aktive-facebook-nutzer-weltweit>

⁴⁴ <https://de.statista.com/statistik/daten/studie/795086/umfrage/anzahl-der-nutzer-von-instagram-weltweit/>

5.5 KI: Gemeingut oder Eigentum?

Immobilien und Gegenstände gelten als Eigentum, Ozeane und die Erdatmosphäre als Gemeingut. Intellektuelle Schöpfungen kennen Graustufen: 70 Jahre nach dem Tod eines Urhebers verfallen Schutzrechte, 20 Jahre nach Patentierung endet die Ausschließlichkeit und das Kartellrecht kennt Pflichten zur Offenlegung von Geschäftsgeheimnissen. Wie könnte nun künstliche Intelligenz einsortiert werden und welche Fragen ergeben sich dazu?

Tragödie der Allmende Die berühmte Tragödie der Allmende beschreibt die Übernutzung von Gemeingut, die den Nutzen für die Allgemeinheit schädigen kann. Sie stellt damit die Frage der *Zulässigkeit*: Sollte es etwa erlaubt sein, Gase in die Atmosphäre auszustoßen und Chemikalien in die Ozeane zu verklappen, und für die Vorteile Einzelner das Gemeingut aller zu gefährden? Ebenso stellt sie die Frage der *Einvernahme*: Sollte beispielsweise einer Firma Micro-Sauerstoff erlaubt sein, die gesamte Atmosphäre einzufangen, sie aufzubereiten, den Sauerstoffgehalt festzulegen, der für uns alle auszureichen habe, sie schließlich in Flaschen zu füllen und die lebensnotwendig gewordene Dienstleistung an uns teuer zu verkaufen? Was soll als Gemeingut und was als Eigentum einzelner verstanden werden und welche Konsequenzen daraus wollen wir zulassen?

Mißbrauch von Gemeingut Die Analogie ist nicht gänzlich überzogen: Bei Suchmaschinen für Handys ist im August 2023 Google mit über 95% Marktanteil unangefochtener Monopolist. Die Firma legt, von sehr wenigen Benachteiligungsverboten abgesehen, die Aufnahme in den Suchindex und die Reihenfolge der Ergebnisse fest und bestimmt damit unsere Weltsicht.⁴⁵

Nutzerverhalten als Gemeingut Ein Beispiel: Wo ist auf Tastaturen das at (@) Zeichen? Windows-Anwender wissen nach tausenden Mails AltGr-Q. Mac-Nutzer wissen das auch: Alt-L. Wo bei Windows die AltGr Taste liegt, positioniert der Mac jedoch die Kommando-Taste ⌘. Anwender, welche die Haptik des AltGR-Q verinnerlicht haben, geben am Mac versehentlich ⌘Q ein, was dort ein Quit ist und die gerade laufende Anwendung schließt. Der häufige Wechsel zwischen Systemen liegt natürlich nicht im Interesse der Hersteller; die grundsätzlich mögliche Veränderung von Tastaturbelegungen ist nicht immer einfach umzusetzen und stellt keine sinnvolle Lösung dar.

Wem gehört eigentlich das in tausenden Bedienvorgängen intuitiv eingeübte Verhalten von Milliarden Anwendern? Wie zulässig soll *digitales nudging*⁴⁶ in der Benutzerschnittstelle sein? Könnte eine Einordnung des von Google erstellten Suchindex oder des von OpenAI erarbeiteten large language models als Gemeingut sinnvoll sein?

Einordnung von KI Wie sind in diesem Raster digitale Dienste einzuordnen, die Wissensbestände erschließen und bewerten, die Fragen beantworten oder Analog-Inhalte erstellen und die dabei an fremden Texten trainiert wurden, oftmals ohne Klärung der Nutzungsberechtigung? Der kollektivistische Ansatz als Gemeingut ist überdenkenswert und könnte neuen Formen von Kartellrecht benötigen, er birgt aber wiederum neue Gefahren. Die Wikipedia realisiert beispielsweise eine solche Allmende von Wissen. Durch die Einfachheit ihrer Nutzung und die hohe Verfügbarkeit von Beiträgen zu fast allen Themen wird sie daher de facto zur einzigen Definitions- und Deutungsinstanz für Begriffe. Dadurch entsteht eine zentrale Verwalterin von Wissen und das gesellschaftliche Verständnis von Wissenschaft wird auf die naive Vorstellung reduziert, es gäbe zu allem die eine, wissenschaftlich korrekte Antwort und diese stünde im Internet.

Die größte Gefahr dieser Vorstellung besteht in der Erosion gesellschaftlicher Diskurse: Es wird nicht mehr diskutiert und abgewogen, untersucht, falsifiziert und validiert, kontextualisiert und beobachtet – sondern es werden autoritative Antworten erwartet, denen eine automatische Richtigkeit unterstellt wird. Wer etwas anderes sagt, als Google, Wikipedia,

⁴⁵Dazu ein persönliches Beispiel: Das für Personen oft angezeigte Google Knowledge Panel zu seinem Namen konnte der Autor zwar mit *Claim Knowledge Panel* „übernehmen“, trotz Feedback und Email-Austausch mit der Hotline gelang es bis heute aber nicht, ein vom Autor beim Springer-Verlag veröffentlichtes und von Google Book indiziertes Werk in die Liste der Bücher des Google Knowledge Panel aufzunehmen. Das entscheide, so die erhaltene Auskunft, der Algorithmus zu gegebener Zeit schon selber.

⁴⁶Man denke beispielsweise an Cookie-Banner, bei denen sich der Surfer für eine Ablehnung der Cookies durch umfangreiche Erläuterungen und Einstellungen durchscrollen muß. Der eine Klick auf die passend hervorgehobene Schaltfläche zur Genehmigung aller Datenübertragungen fällt viel leichter.

ChatGPT oder irgendein anderer digitaler Dienst, ist uninformiert, dumm, manipuliert, lügt oder verbreitet Desinformation. Daher darf ihm mindestens der Schreibzugang zu wirksamen Plattformen der Kommunikation entzogen werden. Die daraus entstehende Intoleranz gegenüber der anderen Meinung ist ein gefährlicher Schritt in Richtung Totalitarismus. Gestern hat die Inquisition 1.0 ihre Gegner am Scheiterhaufen verbrannt – heute nimmt ihnen die Inquisition 2.0 die Plattformen. Die Inquisition 3.0 kann im chinesischen *social score* System besichtigt werden.

Jeder Widerspruch gegen die Vorgehensweise der Datenwissenschaft erscheint als Irrsinn und wird als Ausdruck psychologischer Reaktanz beiseite gewischt. Unangenehme Parallelen drängen sich auf: Die alten Ketzer hatten Gott beleidigt und mußten zur Rettung ihrer Seele im irdischen Feuer gereinigt werden. Die neuen Ketzer beleidigen die Zwangsläufigkeit statistischen Schließens; sie müßten von den Sprachrohren der Gesellschaft fern gehalten werden, da sie diese sonst mit ihren Zweifeln an der rationalen Vorgehensweise verunsichern würden. Die großen Erfolge, mit denen der Verstand viele Irrtümer ausgeräumt hat, machen das Beharren auf fortdauernde Skepsis zu einem glücklosen aber notwendigen Unterfangen. Weitere Kritikpunkte umfassen:

1. Die Aushebelung des Individuums und seines Denkens als der letzten und allein souveränen Instanz der Daseinsdeutung – und damit die Preisgabe eines zentralen Wertes der Aufklärung.
2. Die Reduktion der Welt und des menschlichen Daseins auf das Meßbare – und damit auf das, wozu wir uns heute entschieden haben, es zu messen, auf das, was wir heute tatsächlich messen, unter Ausschluß aller Effekte, die wir heute noch nicht kennen.
3. Die Preisgabe der skeptischen, kritischen Position als Grundlage des Denkens – und damit das Vergessen der historischen Erfahrung, daß sich die allermeisten Hypothesen im Laufe der Geschichte als falsch erwiesen haben und daß ein wesentlicher Wert von Wissenschaft nicht nur aus dem Aufstellen hilfreicher Modellannahmen bestand sondern darin, kostspielige ideologische Irrtümer aufzuklären und zu vermeiden.

5.6 Unvermeidbare Krisen bei technischen Innovationen

Innovationen versprechen Neuerungen, Verbesserungen und Veränderungen. Sie können auch Angst machen, speziell bei Generationen, die für die Herausforderungen ihres Lebens bereits Lösungen gefunden haben. Daher werden Innovationen immer auch mit Skepsis und Gegnerschaft zu kämpfen haben. Diese psychischen und sozialen Prozesse sind oftmals mehr emotional als rational, aber neben der Ratio ist auch die Emotio ein wesentliches Element des menschlichen Daseins. Der Verstand kennt bei der Einführung neuer Technologien nicht alle möglichen Folgen, das Gefühl schwankt daher zwischen einer unreflektierten Begeisterung und den Ängsten vor dem Unbekannten. Vorsicht und Zurückhaltung bei der Einführung von Innovationen könnten daher grundsätzlich angebracht sein.

Innovationen, mit denen Geld oder Macht zu gewinnen ist, werden gerne mit Versprechen eingeführt, welche diese anfängliche Skepsis zerstreuen sollen. Besonders attraktiv erscheint die schnelle Einführung in der Breite, denn wenn jeder ein Handy oder ein Instagram Account nutzt, dann wird das Produkt unabhängig von möglichen Schäden rasch als natürliches oder notwendiges Element des Daseins etabliert. Für die Hersteller, Experten und weiteren Nutznießer der Technologie sichert diese Vorgehensweise Macht, Bedeutung und Gewinn. Häufige Muster sind daher Überschwang und Verharmlosung, gepaart mit Gier und Maßlosigkeit.

Im Ergebnis entstehen spannende Fehl- und Überreaktionen. So wurden Röntgenstrahlen jahrelang bei der Anprobe von Schuhen benutzt, bis die biologischen Risiken der Strahlung erkannt wurden. Andererseits ziehen sich manche Länder aus der äußerst sicheren Kernkraft mit 0,0097 Todesfälle pro erzeugter Terawattstunde Strom zurück, weil ihr Klumpenrisiko psychologisch bedrohlicher wirkt als die Langzeitrisiken der Luftverschmutzung von Braunkohle mit 33 Todesfälle pro erzeugter Terawattstunde Strom [Sov+16; MW07]. Nachdem sie die Problematik des Datenschutzes erkannt hatte, führte die EU ein Datenschutzrecht ein, das uns in der Praxis bei jedem kleinen Blog Cookie-Banner beschert, die verlorene Privatsphäre aber weder bei digitalen Zahlungssystemen, noch bei Patientenakten, Ausweisen oder sozialen Netzwerken repariert.

Bemerkenswert ist auch das Nebeneinander dieser Phänomene in unterschiedlichen Bereichen der Innovation mit zueinander verschobenen Lebenszyklen. So beklagt unsere Gesellschaft

gerade Fehlentwicklungen der Vergangenheit im Energie-, Karbon- und Umweltsektor, führt aber in hoher Geschwindigkeit noch unausgereifte digitale Technologien mit starken gesellschaftlichen Auswirkungen ein, die noch lange nicht wirklich verstanden sind. Zugleich aber zögern manche Funkloch-Länder beim Ausbau der 5G, DSL und Glasfaser-Infrastrukturen in Bereichen, in welchen der Bedarf nach flächendeckender digitaler Kommunikation als unumkehrbare, völlig unbestrittene Anforderung gelten darf. Die Zukunft wird zeigen, ob der AI Act der EU eine bessere Antwort auf KI sein wird als die gut gemeinte DSGVO auf den Verlust von Privatsphäre, Daten- und Prozesssouveränität.

Einerseits müssen neue Technologien rasch in die Praxis getragen werden um dort zu scheitern und an diesem Scheitern weiter zu wachsen. Andererseits birgt ein rasches Ausrollen das Risiko von Schieflagen, die durch Regulierung nur mehr schwer einzufangen ist. Eine zu enge Regulierung wiederum erhöht das Risiko der Staaten, in der technischen und wirtschaftlichen Entwicklung abgehängt zu werden. Insgesamt stellt sich die Situation als ein Konglomerat vieler Widersprüchlichkeiten dar. Kritik fällt daher leicht, eine richtige Antwort erscheint unmöglich und läßt sich nicht auf rationale Weise geben, denn sie hinge nämlich von den Wertesystemen ab, in denen wir leben wollen. Diese jedoch ändern sich, über Generationen, Regionen und stets auch in Reaktion auf die durch uns veränderte Umwelt. Hilfreich könnten daher föderale, regionale und subsidiäre Ansätze sein, die aber im Widerstreit stehen zu globalen digitalen Agenden. Diese begründen sich nicht nur aus monopolaren ideologischen Ansätzen sondern folgen auch aus der Ökonomie des Digitalen, deren Effizienz mit dem Verbreitungsgrad zunimmt und die daher an Interoperabilität, an Standardisierung und möglichst globalem Ausrollen von Systemen interessiert ist. Dieses jedoch steht in Widerspruch zur regionalen Kultur des Menschen und seinem evolutionären Aufwachsen in kleinen Stammesgesellschaften, was sich ethnologisch (Tribalismus) und neurobiologisch (Dunbar Zahl) begründen läßt.

6 Vorstellbare Antworten

In diesem Abschnitt führe ich exemplarisch einige nennenswerte Antworten auf KI an. Meine persönliche Antwort gebe ich mit dem Plädoyer für eine neue Aufklärung am Ende in Abschnitt 8.

6.1 Differenziertere Sichtweise

Die öffentliche Ankündigung von ChatGPT und die Allianz von OpenAI mit Microsoft brachte Bewegung in die öffentlichen Diskurse *und* in die Aktienkurse, mittlerweile wird über AI in vielen Bereichen aufgeregt diskutiert.

Textgenerierende KI ist aber mehr Textgenerator als Intelligenz: Sie liest, modelliert und repetiert. Aus präzise formulierten Texten kann sie mit Glück gute Zusammenfassungen reproduzieren. In Verbindung mit weiteren Werkzeugen, etwa Suchmaschinen oder Rechenprogrammen, kann sie spezifische Anforderungen an diese Hilfsmittel auslagern. Ist textgenerierende KI aber wirklich in einem tieferen Sinne intelligent oder wirkt sie nur dank ihrer Eloquenz so, als ob sie es wäre? PENTLAND schildert⁴⁷, wie sprachliche Fluidität oft als Indikator für bestehendes Expertenwissen gewertet wird, obwohl hier streng genommen kein Zusammenhang besteht. Signale eröffnen stets Chancen für Mimikry, für Selbsttäuschung und für Vortäuschung. Geschieht dieses gerade bei KI?

Auf Instagram findet sich das Foto eines Straßenschilds mit der provokanten Frage: *Before we work on artificial intelligence, why don't we do something about natural stupidity?* Diesen Gedanken kann man weiterspinnen: Wenn KI an menschlichen Texten trainiert wird, reproduziert sie dann nicht diese *natural stupidity*? Wenn sie aber an ihresgleichen trainiert wird, wo kommt dann die menschliche Sinngebung her? Selbst wenn beide Aspekte zufriedenstellend behandelt würden: Welche Bereiche der menschlichen Lebens und unserer Gesellschaft wollen wir einem Textgenerator anvertrauen? Erziehung, Politik oder gar Entscheidungen? Die Diskussion wäre weniger aufgeregt, wenn wir textgenerierende KI nur als schlichtes Hilfsmittel betrachten würden.

Gänzlich andere Fragen stellen sich natürlich, sobald wir Algorithmen mit Aktoren verbinden und in die Welt entlassen, etwa als Roboter-Soldaten oder Computer-Polizisten mit Waffen

⁴⁷in <https://www.youtube.com/watch?v=UL31rrUXFIk>

und Entscheidungsbefugnis. Mit Hinblick auf die alte lateinische Erfahrung *si vis pacem para bellum*⁴⁸ muß dieser Irrsinn gleichwohl untersucht werden, vermutlich wäre aber eine öffentliche Ächtung des Einsatzes sinnvoll wie auch Regelungen, die den Nichtweitergabe-Vereinbarungen bei anderen Massenvernichtungswaffen entsprechen.

Ein Teil dieser differenzierteren Sichtweise könnte also das unaufgeregte Eingeständnis sein, daß Systeme wie ChatGPT mindestens auf absehbare Zeit keine Intelligenzen darstellen, wenn gleich sie auch hilfreiche Suchknechte sein können.

6.2 All Watched Over by Machines of Loving Grace

Sicherheit und Geborgenheit? Bieten uns Maschinen vielleicht doch Sicherheit und Geborgenheit? Das berühmte Gedicht *All Watched Over by Machines of Loving Grace* von R. BRAUTIGAN⁴⁹ greift diesen Gedanken auf, doch scheint nicht geklärt, ob der Autor seine Zeilen als Ironie oder als Hoffnung gemeint hat.⁵⁰ Die gleichnamige Dokumentation von A. CURTIS kommt jedenfalls zum Schluß, daß *Computer die Menschheit nicht befreit haben sondern unser Bild von der Welt um uns herum verzerrt und simplifiziert haben.*⁵¹

<p>I like to think (and the sooner the better!) where mammals and computers live together in mutually programming harmony like pure water touching clear sky.</p>	<p>I like to think (right now, please!) of a cybernetic forest filled with pines and electronics past computers as if they were flowers with spinning blossoms.</p>	<p>I like to think (it has to be!) of a cybernetic ecology where we are free of our labors and joined back to nature, returned to our mammal bothers and sisters, and all watched over by machines of loving grace.</p>
---	---	---

Die Frage, wie lange es noch braucht, bis eine allgemeine künstliche Intelligenz entwickelt ist, die alle kognitiven Fähigkeiten eines Menschen hat, und ob es eine solche überhaupt jemals geben kann, wird unter Experten unterschiedlich diskutiert. Es liegt, jedenfalls derzeit noch, in unserer Hand, ob wir die angebliche (!) Sicherheit einer solchen Welt anstreben oder sie als eine Dystopie fürchten, die noch den letzten Rest menschlicher Freiheit vernichtet.

Dystopien vermeintlichen Glücks Ein Blick in literarische Vorlagen kann die Problematik dieses Prozesses verdeutlichen, denn die Katastrophen vieler Dystopien sind nicht aufzuhalten: Sie bestehen ja gerade darin, daß ihre Einwohner sich glücklich wähnen, nicht deshalb, weil sie tatsächlich glückliche Menschen wären, sondern weil die Entwicklung in die Dystopie die Voraussetzungen dafür schafft, das Ergebnis als glücklich zu empfinden.

In HUXLEYS *Brave New World* werden Gefühlsschwankungen durch eine Droge namens Soma ausgeglichen und eine Sexualität, die gesellschaftlich auf hemmungsloses Ausleben angelegt ist, läßt tiefere Emotionen gar nicht erst aufkommen. Der Film *Matrix* geht noch weiter: Maschinen nutzen die Körper der Menschen zur Gewinnung von Energie und halten diese künstlich ernährt in einem Hochregallager unter Kontrolle. Dort sind sie neurologisch mit einer simulierten Scheinwelt verbunden. In beiden Dystopien sind fast alle Menschen „glücklich“, nur die Hauptfiguren leiden: Neo, der sich entschließt der Simulation der Matrix zu entkommen, und bei HUXLEY der Wilde, der seine Werte aus einer Gesamtausgabe von SHAKESPEARE bezieht und sich in der Welt von Drogen, Spiel und Spaß sehr unwohl fühlt.

Wir können das Bild zurück übertragen, in die KI und weiter in die Digitalisierung. Viele Leiden und Probleme der digitalen Gesellschaft sind uns bekannt. Gleichwohl sind wir nicht in der Lage, sie als Bedrohung zu erkennen, da wir gerade hektisch damit beschäftigt sind, sie unwiderrufflich als Teil einer unverzichtbaren Infrastruktur zu etablieren.

Digitale soziale Netze liefern bei der Interaktion intensive Ausschüttungen von Dopamin und stellen ein neues Suchtphänomen dar, trotzdem – oder eher: deshalb – wachsen Facebook und Instagram immer weiter. Diese Dienste stellen das HUXLEYSche Soma der digitalen Umwelt dar. Sie fabrizieren das Paradoxon, daß wir über sie weltweit Kontakte schließen und über Entfernungen halten, Psychologen aber zugleich die wachsende Einsamkeit als Megatrend

⁴⁸ Wenn Du Frieden willst dann rüste für den Krieg.

⁴⁹ Zitiert nach <https://web.archive.org/web/20211125232026/http://www.brautigian.net/graphics/machines/machines-loudspeaker.gif>, Schreibweisen aus dem Original übernommen.

⁵⁰ https://en.wikipedia.org/wiki/All_Watched_Over_by_Machines_of_Loving_Grace

⁵¹ [https://en.wikipedia.org/wiki/All_Watched_Over_by_Machines_of_Loving_Grace_\(TV_series\)](https://en.wikipedia.org/wiki/All_Watched_Over_by_Machines_of_Loving_Grace_(TV_series)).

des 21. Jahrhunderts beklagen. Menschlicher Kontakt kann über Touchscreens nicht *wirklich* vermittelt werden, daher müssen wir lernen, mit den Surrogaten glücklich zu werden.

Ähnliche Paradoxien bestehen auch bei texterzeugenden Maschinen. Sie rezipieren und rezipieren textuelle Wahrnehmungen und werden zum Echoraum der Masse. Sie können den Einzelnen noch überraschen, weil er nicht alle Texte gelesen hat, bieten aber insgesamt keine neuen Empfindungen oder Erkenntnisse. *Echte* künstliche Intelligenz könnte mit Problemlösungen aufwarten, aber sie verflacht das Leben. Der Roman *Transfer* von STANISŁAW LEM beschreibt die Entfremdung von Astronauten, die nach 10-jähriger Entdeckungsreise in eine Gesellschaft zurückkehren, die aufgrund der relativistischen Zeitdilatation inzwischen 127 Jahre gealtert ist: Langweilig erscheint sie ihnen, übersättigt und ohne jede Neugierde. Viele können sich nicht integrieren und sie planen einen neuen Flug ins Abenteuer. HUXLEY und LEM zeigen, wie Zeitgenossen die langsamen Veränderungen in die Dystopie nicht erkennen und nur die Konfrontation mit dem gänzlich Anderen den Abgrund sichtbar macht. Das Glück dieser Welten ist unaufhaltbar aber zweifelhaft.

6.3 Die Wiederentdeckung des Menschlichen

Eine mögliche Gegenposition zu diesem zweifelhaften Glück kann die bewusste Rückbesinnung sein auf jene Aspekte, in denen der Mensch besser ist und wohl auch immer besser sein wird als die Maschine. Die Kunst wird darin bestehen, herauszuschälen, was eben dieses *Menschliche* genau ist. Die Debatte über die KI können wir als Ordnungsruf verstehen, darüber nachzudenken.

Bald nach der Vorstellung von ChatGPT formulierte ich auf LinkedIn einen kurzen Text, der mir mittlerweile nicht mehr so naiv erscheint:⁵²

Schon der Fernseher lehrte mich den Duft von Blumen und Parfum,
Google lehrte mich den Unterschied zwischen Texten und Wissen,
und Twitter die Freude an ausführlichen Beschreibungen.
Facebook lehrt mich, nur aufrichtige Komplimente zu machen,
und Zoom den kräftigen Handschlag zur Begrüßung des Geschäftspartners.
Das Handy schließlich lehrt mich die Umarmung beim Abschied der Freunde.
Maschinen lehren mich die Bedeutung des Menschlichen!
ChatGPT nun macht mich schlauer!
Es lehrt mich den Unterschied zwischen Ausbildung und Bildung,
es hilft mir, die Bedeutung des menschlichen Lehrers zu würdigen
und seine Geduld zu schätzen,
es gemahnt mich, die Zuverlässigkeit von Antworten zu prüfen,
und es erinnert mich an die Freude des Findens,
wenn ich selber mühsam nach den rechten Worten gesucht habe.
Danke ChatGPT!
Im Zeitalter der Digitalisierung lerne ich von Dir,
Mensch zu sein, die Maschinen unter den Menschen besser zu erkennen
und die Begegnung mit ihnen zu vermeiden.

ADALBERT STIFTER schildert in seinem Roman *Der Nachsommer*, wie sich nach der Revolution von 1848 der Freiherr von Riesach aus seinen Ämtern zurück zieht, weil er keine Möglichkeit sieht, das in ihnen zu realisieren, was er für notwendig hält. Der Maßstab, den er dabei verfolgt, ist *die Ehrfurcht vor den Dingen, wie sie an sich sind, und das, was die Dinge nur für sich forderten, und was ihrer Wesenheit gemäß war*. Eine von Ideologien getriebene und von Eigeninteressen geleitete Welt verfehlt dieses Ziel. Er renoviert den alten Aspermeierhof und schafft dort ein *gesegnetes und von Gott beglücktes Land*. Er achtet in Filzpantoffeln die kunstvollen Intarsien im Holzboden der Bibliothek ebenso wie die Vögel und Insekten des Anwesens und die Arbeiter auf seinem Hof. Regelmäßig läßt er sie die Stämme der Birkenbäume waschen, denn er weiß, daß *den Wert von Pflanzen keiner vollständig ermessen kann, als der sie pflegt. Jedes Wissen hat Ausläufe*, so ist ihm bewußt, *die man oft nicht ahnt und man soll die kleinsten Dinge nicht vernachlässigen, wenn man auch noch nicht weiß, wie sie mit den größeren zusammenhängen*.

Sind uns in einer Zeit kürzester und schnellster Nachrichten – 160 Zeichen SMS und 280 Zeichen Twitter – die kulturellen Welten aus Büchern von über 800 Seiten noch zugänglich?

⁵²https://www.linkedin.com/posts/clemens-h-cap-489691257_schon-der-fernseher-lehrte-mich-den-duft-activity-7045716130920820736-jKMe/

Oder liegt, gerade umgekehrt, genau in ihnen die Chance einer Wiederentdeckung vergessener Werte? Natürlich schildert STIFTER hier Ideale, die er zur Abgrenzung von der Zeit des Vormärz zusätzlich noch literarisch übersteigert. Diese Vorstellungen von einer *Ehrfurcht vor den Dingen und ihrer Wesenheit* sind einer 4. industriellen Revolution entglitten und diesen Verlust muß man nicht unbedingt als Fortschritt empfinden oder verkaufen.

Die Gefahren der KI liegen nicht in ihren Fähigkeiten oder ihren Fehlern. Das Problem ist die Bereitschaft vieler Menschen, ihre Lebenswirklichkeit auf das zu reduzieren, was durch Geräte gemessen, durch Wissenschaft formal beschrieben und von Computern symbolisch verarbeitet werden kann. Wer wie viele *digital natives* eine andere Welt nicht kennt – wie sollte er das Unbekannte vermissen? Wer aber dennoch eine andere Welt sehen will wird sie in den digitalen Quellen nicht finden. Der Kreis schließt sich.

Doch ist dieser Wandel der Welt auf *evolutionäre* Weise umkehrbar? Die Gläubigkeit an technokratische Lösungen und die Abhängigkeit unserer Infrastrukturen von diesen Lösungen erlauben nur mikroskopische Veränderungen ohne eine massive Beschädigung der Gesellschaft. Technologische Entwicklungen haben historisch oft *revolutionäre und blutige* Ereignisse verursacht.⁵³ Wer den historischen Blutzoll solcher Umbrüche kennt, kann sie sich nicht wünschen, und wer die Effizienz der digitalen Überwachung kennt, hält sie für nicht realisierbar.

7 Die trügerische Sehnsucht nach dem Orakel

Die Unsicherheit menschlicher Lebenserfahrung weckt die Hoffnung nach *einem, der weiß wie es wird und geht*. Zunächst waren das die Götter. Sie lebten in Familienstruktur, im Olymp oder auf Walhall, und griffen gelegentlich in das menschliche Schicksal ein, weshalb wir uns mit ihnen gut stellen wollten. Dazu gab es einen eigenen Berufsstand, der uns die notwendigen Rituale zur Besänftigung erklärte. Deren Einhaltung war dann eine Frage der Überzeugungskraft dieses Berufsstandes und der sich einstellenden Erfolge. Die Religionswissenschaften kennen Gebete, aber auch Tier- und Menschenopfer sind nachweisbar. Aufgrund der geringen Auswirkungen auf die Ernte schwand schließlich die Akzeptanz und wir entdeckten andere Wege, um die Gewalten der Natur zu erklären und unsere Lebenssituationen zu verbessern. Macht und Einfluß verschoben sich zu jenen Berufen, welche die neu entdeckten Wahrheiten und Wirklichkeiten angeblich besser deuten und verwalten zu können glaubten.

Rezeption der Narrative Das grundsätzliche gesellschaftliche Spannungsfeld jedoch blieb erhalten: *Einerseits* die Überzeugung Einzelner, was zu tun sei, die sich wahlweise auf Gebete, Eingebungen, Experimente, Hypothesen oder andere Methoden gründete, *andererseits* die Bereitschaft der Vielen, diesen Überlegungen zu folgen, und die sich mangels eigener Kontakte zum jeweiligen Orakel aus anderen Quellen speisen mußte: Predigten und Vorlesungen, schließlich Lehrbücher und Massenmedien. Mit dem Aufkommen der Buchkultur wurden die Verbindungen zur Weisheitsquelle besser. Internet und digitale soziale Medien verschoben diese Mechanismen wieder, in Richtung auf eine spontanere und emotionalere Ausbildung von Dissens und Konsens.

Die höhere Entwicklungsgeschwindigkeit eng gekoppelter, globaler, technologisch fortgeschrittener Gesellschaften erhöht die Notwendigkeit rascher, koordinierter Reaktionen auf neue Situationen und mit ihr auch das Bedürfnis, schneller einen gesellschaftlichen Konsens (N. CHOMSKY) herzustellen, der zur Umsetzung von Entscheidungen notwendig erscheint. Dafür sind Überzeugungs- und Rechtfertigungsinstanzen notwendig.

Instanzen der Rechtfertigung Die Erkenntnistheoretiker argumentieren⁵⁴, daß Naturwissenschaften weder Wahrheiten noch Gewißheiten erzeugen und vor allem darin gut sind, Fehler in Modellen nachzuweisen. Gleichwohl lassen sich Wissenschaft und die mit ihr gerne verbundenen Begriffe von *Wirklichkeit* und *Wahrheit*, trefflich als Rechtfertigungsinstanzen

⁵³Die Erfindung von Faustfeuerwaffen und Kanonen machte MAXIMILIAN I zum letzten Ritter und markierten den Beginn der Neuzeit, die Erfindung des Buchdrucks wird gerne mit der französischen Revolution und der Reformation verbunden, die industrielle Revolution im 18. Jahrhundert beschleunigte den Imperialismus, verursachte Arbeiteraufstände und soziale Unruhen, die sozialen Medien schließlich spielten eine wichtige Rolle im sogenannten arabischen Frühling.

⁵⁴Insbesondere K. POPPER [Pop89] und W. STEGMÜLLER [Ste69].

einsetzen. *Science* wandelt sich, beschleunigt durch gesellschaftliche Krisenwahrnehmung, zu *Scientismus*, also zur Überzeugung, daß alle Probleme wissenschaftlich behandelt werden könnten und sollten. Entscheidungsträger, welche ihre Entscheidungen nicht mehr *tragen* können oder wollen, berufen sich auf „die Wissenschaft“. Aktivisten rufen „follow the science“ und prägen das Schlagwort „Science for Future“. Durch diese Entwicklungen wird eine neue Sicht auf Wissenschaft erzeugt, die endgültige Antworten und damit verbundene Sicherheiten verspricht.

Populistische Positionen arbeiten gerne mit einfachen, sprachlich wie argumentativ leicht zu transportierenden Bildern. Sie benötigen aber Instanzen, die bereit sind, ihre Korrektheit zu bestätigen. Oft genug tat Wissenschaft der Politik diesen Gefallen, indem sie Sicherheiten vorgab, obwohl ihre Geschichte eher von einer Reihe erfolgreich aufgefundener Irrtümer berichtet – was im Nachhinein verdrängt wird und durch einen *survivor bias* verloren geht.

Mechanismen der Korrektheit waren durch lange Zeit die eigenen Wiederholungen von Experimenten oder das eigene Nachvollziehen von Gedanken. Mit der Komplexität der Disziplinen und den gestiegenen Kosten für Apparaturen wandelten sich die Mechanismen zu Referenz, Zitat und schließlich Konsens. Als wahr gilt nun, was anerkannt und zitierfähig ist, die Verbindung zum erkenntnistheoretischen Urgrund geht dabei immer mehr verloren.

KI als wahrgenommene Instanz der Rechtfertigung Nun läuft KI der Wissenschaft den Rang ab. Generative KI kann überzeugendere Texte schneller formulieren, stützt sich auf statistische Beobachtungen und hat einen größeren Korpus bestehender und daher sozial erfolgreich rezipierter Texte bearbeitet, als es ein Wissenschaftler in seinem ganzen Leben je könnte. In der sprachlichen Argumentation und ihrer sozialen Wirkung wird KI bald nicht mehr zu schlagen sein.

ChatGPT aber denkt nicht systematisch und scheitert daher bereits an einfachsten Aufgaben, wenn es sie nicht aus dem Textkorpus kennt und Antworten über sprachliche Analogstrukturen herleiten kann. ChatGPT geht nicht ins Labor, es macht keine Experimente und es versucht auch niemals, Thesen zu falsifizieren. Generative KI bleibt gleichwohl ein hilfreiches Werkzeug, weil es schneller lesen und rascher zusammenfassen kann, als der Mensch. Diese Fähigkeit macht das System zu einer sozial wirksamen *Rechtfertigungsmaschine* ohne echte Verbindung zu anerkannten Mechanismen der Erkenntnis.

Ermächtigung durch Wahrheit Ein weiteres Problem könnte Sorgen bereiten: *Wer überzeugt ist, über die richtigen Antworten zu verfügen, so meinte sinngemäß⁵⁵ der Philosoph HERMANN LÜBBE, wird sich früher oder später berechtigt fühlen, Gewalt auszuüben.* Diese Selbstermächtigung zur Gewalt muß sich nicht auf körperliche Gewalt ausdehnen, sie kann auch die Übernahme und Einübung jener Werte und Verhaltensweisen verlangen, die man selber für richtig hält und daher von den anderen übernommen sehen will.

KI könnte zukünftigen Entscheidern diese Überzeugung vermitteln, sie würden über die richtigen Antworten verfügen. Sie stützt sich methodisch auf die datentechnische Erfassung immer größerer Ausschnitte der Welt und die Ableitung der statistisch bestmöglichen Modelle. Wer wollte es schon wagen, gegen empirisch gewonnene Fakten oder daraus destillierte Wahrheiten und Wirklichkeiten zu argumentieren?

Hier ist der Philosoph KONRAD LIESSMANN ein engagierter Kritiker [Lie17]: *Empirie, so formulierte es einmal mit unangenehmer Schärfe der Philosoph Günther Anders, ist nur etwas für Idioten. Denen mangelt es nämlich an der Fähigkeit, über das Handgreifliche hinaus zu denken.* Die eingeschränkte Weltsicht der KI, ihre Präkonditionierung durch die vorgegebene Architektur, die Auswahl des Trainingskorpus und die Nachjustierung sind bereits gravierende Einschränkungen. Zudem müssen wir anerkennen, daß viele gedankliche Modelle, beispielsweise der modernen Physik, so sehr sie auch mit empirischen Beobachtungen verbunden sind, alles andere als handgreiflich naheliegend sind – sie gehen weit über eine reine Modellierung experimenteller Tatsachen hinaus.

Selbst wenn man diese Argumente beiseite wischt: So ehrenhaft die Motivation und so solide die Ermächtigung auch sein mag, [d]er Versuch, den Himmel auf Erden einzurichten erzeugt stets die Hölle [...] führt zu Intoleranz, zu religiösen Kriegen und zur Rettung der Seelen durch die Inquisition [Pop92]. Der Schriftsteller C. S. LEWIS meinte: *Of all tyrannies, a tyranny sincerely exercised for the good of its victims may be the most oppressive.* [Lew79]

⁵⁵Hier zitiert nach einer Formulierung von PAUL WATZLAWICK <https://www.youtube.com/watch?v=c14aZTPsTSs> 39:30, siehe auch [Lüb89], S. 120.

Die Botschaften, mit denen KI und allgemeiner digitale Technologien und Transformationen beworben werden, versprechen gerne diesen Himmel auf Erden. Suchmaschinen und text-generierende KI nehmen zentralen Einfluß auf unsere Weltwahrnehmung. Die eine Technik *ist* und die andere *wird* unvermeidbar für unser Überleben. Sie werden damit unsere Welt. *Schäden* gehen nach dieser Überlegung nicht von der Technologie selbst aus, sondern von ihrer Wirkung auf die menschliche Gesellschaft und ihre Entscheidungsstrukturen.

Fehlende Demut und Skepsis Wissenschaft ist unbestreitbar ein wesentlicher Motor des menschlichen Fortschritts geworden. Dieser Erfolg hat, mit gutem Grund, eine wissenschaftliche, rationale Weltsicht als die scheinbar „richtige“ etabliert.

Trotz dieser wundervollen Erfolge kann sie aber nicht mit tieferen Begründungen und Evidenzen aufwarten, denn unsere *Entscheidungen* hängen weiterhin von subjektiven Werten und individuellen Zielen ab. Eine durchgängig skeptische und demütige Perspektive bleibt auch schon deshalb angebracht, weil sich wissenschaftliche Theorien immer nur auf etwas vorläufig Erkanntes erstrecken. Diese subjektiven Wertungen finden sich auch in den Trainingsdaten der KI und werden von dieser reproduziert. Dieses Problem der Subjektivität läßt sich nicht reparieren, indem man die Trainingsdaten filtert oder einen vermeintlichen Bias der Algorithmen nachkorrigiert, denn die hochmütige Hoffnung, es gäbe den einen *richtigen* Bias ist ja gerade das Problem.

Die Erfolge einer wissenschaftlich-technischen Weltsicht, verbunden mit menschlicher Maßlosigkeit und Gier, können gravierende Probleme verursachen, von denen genügend in der heutigen Welt zu finden sind. Die selbstkritische Weltsicht allerdings, welche die Wissenschaften ursprünglich groß gemacht hat, scheint ihr in einer Phase kurzfristiger technologischer Begeisterung abhanden gekommen zu sein.

Die Ursache des Problems liegt nicht in den beeindruckenden Artefakten, die Wissenschaft und Technik seit der Mitte des vergangenen Jahrhunderts geschaffen haben. Wir finden sie eher in der menschlichen Gier und Ungeduld bei der ökonomischen und politischen Auswertung dieser Entwicklung und in der Unfähigkeit der Massengesellschaft, dem Leben menschlich, mit Herz und Empathie entgegenzutreten. Kann uns KI hier helfen? Hilft sie uns, wenn sie, von einem Textkorpus trainiert, am Fließband bukolische Gedichte für uns schreibt? Hilft sie uns, wenn sie mit dem Argument des bestmöglichen statistischen Modells und der bestmöglichen Antwort die verbleibenden Unsicherheiten des Lebens zu übertünchen versucht?

Der Stolz auf das Denken scheint angesichts vieler Erfolge gerechtfertigt, doch bleibt er nicht die einzige menschliche Fähigkeit. Es zeichnet den Menschen meiner Meinung nach eine Hoffnung aus, mit jeder schwierigen Situation, die er nicht handhaben kann, gleichwohl umgehen zu lernen und damit einen tieferen, schicksalhaften Sinn in sie hinein zu legen, den sie vorher, ohne die individuelle Bewältigung nicht hatte. Viele Gedichte, Gemälde und Musikstücke zeugen von dieser Fähigkeit ebenso wie die gedankliche Befassung mit dem Tod. Ein Zeitgeist allerdings, dem die Hoffnung im Scheitern abhanden gekommen ist, vermeidet die Macht dieser tiefen Emotionen. Fast alles hat als machbar zu gelten und muß *mindestens* systematisch optimiert werden – dem verbliebenen Rest an Unsicherheit weichen wir panisch aus.

8 Zur Notwendigkeit einer neuen Aufklärung

Aufklärung kann als Wandel menschlicher Lebenseinstellungen unter äußeren Einflüssen verstanden werden. Sie ist damit eng mit technologischen Veränderungen verbunden. In der Tradition der Numerierung von Trends können wir drei Wellen unterscheiden.

Aufklärung 1.0 bedeutete in diesem Sinne die Befreiung des Menschen von ihrer Bevormundung durch Autoritäten und deren Verkörperungen in Göttern, Herrschern und heiligen Büchern. Technologisch war sie unter anderem durch den Buchdruck getrieben und die resultierende Verfügbarkeit von Wissen für breitere Bevölkerungsschichten. Ihr von KANT formulierter Leitspruch war: „Habe Mut, Dich Deines *eigenen* Verstandes *ohne Leitung durch einen anderen* zu bedienen.“ Ihr Problem wurde die Überbetonung der Ratio, die durch die Romantik sowie den Sturm und Drang aufgegriffen wurde.

Weitere Veränderungen folgten nun dem technologischen Fortschritt: GOEBBELS Volks-

empfänger, der Fernseher und die Theorie der *public relations* (P. BERNAYS), das globale Dorf (M. McLuhan) und die Massenmedien als einzige wesentliche Quelle von Informationen über die Welt (N. Luhmann) markieren Einschnitte und Probleme. Schließlich entfiel mit dem Internet die früher aus ökonomischen Überlegungen noch erforderliche Prüfung der Inhalte auf Richtigkeit und Wirkung vor ihrer Verteilung.

Zunächst gab es eine Kontraktion auf wenige, zentralisierte Quellen von Meinungen, ein Mechanismus, den N. Chomsky als gesellschaftlich notwendiges System zum Herstellen von Konsens beschrieb (*manufacturing consensus*). Dann wurde mit dem Internet die Kommunikation schneller und nochmals kostengünstiger. Sie eignete sich zunächst zur Ergänzung und Korrektur zentraler Meinungsmonopole, etwa in Usenet, Blogs und alternativen Medien, und wurde dann rasch zu einem politischen und ökonomischen Instrument ausgebaut. Dieses wurde in der Politik zur Wählerwerbung (Obama), dann mittels Microtargetting zur gezielten Wähler-Beeinflussung (Trump, Cambridge Analytica) genutzt und dient mittlerweile auch der systematischen politischen Desinformation. Der *digitale Feudalismus* kommunikativer Plattformen [Cap17], [Cap19] macht eine neue Aufklärung notwendig. Das Schlagwort der Aufklärung 2.0 kennt bereits seine Aktionstage.⁵⁶

Aufklärung 2.0 bedeutet die Befreiung des Menschen von der Bevormundung durch die Massen und ihre Verkörperung in sozialen Netzwerken. Diese haben über die Kommunikationstechnologie neue und globale Wirkmechanismen erlangt. Die Auseinandersetzungen betreffen ideologisch geprägte Mehrheitsbewegungen und sie kennen das Spannungsfeld von *shitstorm* und *deplatforming*. Als Leitspruch könnte eine Aufklärung 2.0 formulieren: Habe Mut zu einer *anderen* Meinung als Deine digitale Blase, *prüfe* sie laufend und sei *bereit*, sie jederzeit aufgrund neuer Beobachtungen zu *verändern*.

Die technologische Entwicklung ist so rasant, daß bereits eine weitere Version von Aufklärung notwendig erscheint.

Aufklärung 3.0 bedeutet die Befreiung des Menschen von der Bevormundung durch digitale Systeme und ihren Ausprägungen als digitales Nudging, smarte Assistenz und künstliche Intelligenz.

Als Leitspruch könnte eine Aufklärung 3.0 formulieren: Habe Mut zur *eigenen Gestaltung* Deines Lebens, auch wenn diese im Widerspruch zu den digitalen Artefakten in Deiner Umgebung steht. Erkenne, daß die Nützlichkeit von Inhalten nicht allein in ihrer Existenz, Herkunft oder postulierten Wahrheit begründet liegt, sondern in *Deiner eigenen Entscheidung* über sie und in Deiner Bereitschaft und Fähigkeit, für diese die Verantwortung zu tragen. Schärfe Deine Urteilskraft! Nutze digitale Werkzeuge für eine Verbesserung Deines Lebens, statt Dich zum Objekt ihrer Manipulation und Beeinflussung zu machen.

Dieser Leitspruch einer Aufklärung 3.0 ist der Gegenpol zu dem, was ich ironisch als das traurige Mantra der künstlichen Intelligenz und eines digitalen Paternalismus wahrnehme und so schwarzmalen würde:

Habe Mut,
gedankenlos Deiner eigenen KI zu folgen,
denn sie ist der Ausgang des Menschen
aus seiner unverschuldeten Freiheit
in die Geborgenheit der Bevormundung
durch schlaue Maschinen.

Abhängigkeit von Wissen und Werkzeug: Kant versprach mit der Aufklärung die Befreiung des Menschen aus einer geistigen Abhängigkeit. In der praktischen Anwendung ist die geforderte Nutzung des eigenen Verstandes aber von Informationen über die Welt abhängig. Die Anfänge der Aufklärung gewannen dieses Wissen durch eigene Beobachtung und aus der sich entwickelnden Buchkultur und ihren Enzyklopädien, bis N. Luhmann darauf hin wies, wie wir fast alles, was wir von der Welt wissen, aus Medien beziehen. In diesen stehen nun zwei Kräfte einander gegenüber: Eine *Zentralisierung*, die durch den Preisverfall der digitalen Informationserstellung und -verbreitung sowie durch eine Gratis-Mentalität im Internet befördert wird und die sich auch in der Krise des Journalismus und der Printmedien

⁵⁶<https://vdw-ev.de/portfolio/aktionstage-aufklaerung-2-0/> und <https://www.youtube.com/watch?v=hvj6BNMACmg>

zeigt – und eine *Dezentralisierung*, die den Bezug einer Vielzahl von Ansichten und Korrekturen aus Blogs, aus Augenzeugenberichten und geleakten Informationen ermöglichen. Vorfälle, die mit den Namen WIKILEAKS, J. ASSANGE oder E. SNOWDEN verknüpft sind, entreißen den politischen Akteuren die Darstellungs- und Deutungsmacht über die verbreiteten Narrative. Diesen ist das ohne Zweifel unangenehm – was aber macht diese Entwicklung mit der übrigen Gesellschaft?

Desinformation von subversiven Kräften aber auch von staatlichen Akteuren spitzen den Konflikt weiter zu, da sie Regulierung unvermeidbar erscheinen lassen. Fact Checking ist nur eine vordergründige Lösung, denn die bereits vor über 2000 Jahren vom lateinischen Satiriker JUVENAL gestellte Frage *Wer bewacht die Wächter?* darf weiterhin als ungelöst gelten. Fact Checker können sich irren bis sie schließlich selber als Desinformatoren auftreten. Mit den deep fake Technologien der bild- und tongenerierenden künstlichen Intelligenz schwindet die Überzeugungskraft von Dokumenten, deren Authentizität nicht mehr zuverlässig prüfbar ist.

Validierung durch eigene Beobachtung oder eigene Urteilskraft hat sich zu einer Art *Pseudovalidierung* durch Referenz gewandelt. Wie aber könnte ein Hinweis auf Dokumente für deren vermeintliche „Richtigkeit“, Nützlichkeit oder Anwendbarkeit bürgen? Validierung durch Masse wäre denkbar: Was viele glauben *könnte* wenn nicht wahr so zumindest hilfreich sein. So lange Information verkörpert weitergegeben wurde, als Buch, das für Satz, Druck und Verteilung gewisse Kosten verursachte, konnte man sich ein wenig auf die bereinigenden Marktmechanismen von Angebot und Nachfrage verlassen. Das Argument der Zahlungsbereitschaft Dritter ist mit den fast völlig verschwundenen Kosten der Informationsverteilung aber entfallen und hat sich in sein Gegenteil gekehrt: Informationen, für deren Verbreitung Dritte bezahlen, zielen oftmals auf Beeinflussung.

Als soziale Wesen fühlen wir uns abhängig von den Einflüssen der Gemeinschaft: Wenn viele einer bestimmten Meinung sind, so fällt es uns schwer, Entscheidungen zu treffen, die von der wahrgenommenen Mehrheitsmeinung abweichen.⁵⁷ Mechanismen der Validierung durch Überlegung und Beobachtung weichen einem Wettstreit um Aufmerksamkeit. In den digitalen sozialen Medien wird diese in den Zahlen von *Clicks, Likes* oder *Followers* gemessen.

KI und Gegenaufklärung ADORNO und HORKHEIMER berichten, wie die Aufklärung keine Freiheit geschaffen hat, sondern nur neue Formen der Macht [HA20]. In einer digitalisierten sozialen Welt ist diese Macht nun viel leichter zu verwalten und setzt daher ihre Schaltstellen mindestens konspirativen Gerüchten aus. Künstliche Intelligenz ist die logische Absetzbewegung von dieser Kritik und kann daher als das *trojanische Pferd der Gegenaufklärung* angesehen werden: Vordergründig nützlich, langfristig aber eine Entmündigung, deren neue Abhängigkeiten mindestens *willkommene Kollateralschäden einer neuen Bequemlichkeit* darstellen.

JAMES BRIDLE bringt es in *New Dark Age* noch drastischer auf den Punkt [Bri18]: Eine technologisch unterstützte Aufklärung könnte das Gegenteil ihrer Versprechen bewirken: Die Unsicherheit einer nicht mehr handhabbaren Komplexität, die sich in Algorithmen, Überwachung und der Aushöhlung von Empathie niederschlägt. Weil wir die Welt um uns nicht mehr verstehen suchen wir unsere Zuflucht im Irrglauben, daß die systematische Analyse von Daten ein stabiles, sicheres und konsistentes Modell unserer Welt ergeben würde.⁵⁸ Künstliche Intelligenz beschert uns damit einen *Neuen Gott* namens ChatGPT, den wir verehren und anbeten sollen [Cap23], der aber nur eine Neuauflage überwunden geglaubter Heilslehren und Ideologien darstellt. KI begründet sich stabiler als zuvor nicht über Höllendrohungen und dunkle Mächte sondern über eine angebliche Rationalität, deren Heilsversprechungen gerne akzeptiert werden, weil die Diskussion ihrer Fragilität sonst die erwünschte Sicherheit in Frage stellt.

Das Glück der Masse In einer größeren Gesellschaft von Menschen, die in einem vordergründig nicht-autoritären System zusammenleben wollen, und in der sich Regelungsmechanismen nach Mehrheiten demokratisch organisieren, *muß* die anonyme, gesichtslose Masse

⁵⁷Vergleiche die sozialpsychologischen Experimente von S. ASCH zur Konformität, von S. MILGRAM zur Gehorsamkeit, das Robbers Cave Experiment von M. SHERIF zur verbindenden Kraft übergeordneter Narrative, oder das Stanford Prison Experiment von P. ZIMBARDO über die Kraft der Stereotypen und Erwartungen.

⁵⁸Er beschreibt die Gefahr als [...] *solutionism: The belief that any given problem can be solved by the application of computation. Whatever the practical or social problem we face, there is an app for it.*

glücklich sein und *nicht* die Ansammlung der vielen Individuen, die allenfalls Schutzpositionen für Minderheitsperspektiven erarbeiten können. Statistische Werkzeuge, Algorithmen und KI sind geeignete Hilfsmittel zum Erwerb und zum Erhalt von Macht in solchen Systemen. Das tiefere Problem ist dabei nicht die KI, sondern die Bereitschaft der vielen Individuen, die Wahrheits-, Bewertungs- und Deutungshoheiten über ihr Leben *ganz bequem und höchst effizient* an eine Maschine zu delegieren.

Diese Welt wird kein Schreckensszenario sein – jedenfalls nicht für die durchschnittliche Mehrheit, auf deren *wahrgenommene* Zufriedenheit und Satttheit keine Gesellschaft langfristig verzichten kann. Der „Wohlfühlstaat“ wird seine häßliche Fratze aus Überwachung und Folter nur jenen Kritikern zeigen, die ihn hinterfragen. Der Einwohner der *Brave New World* bei HUXLEY wird zufrieden vom Wohlfühlkino zur Sexorgie pilgern, im Zweifel kann er sich mit Soma beruhigen, schlimmstensfalls wird er ruhiggestellt. Leiden wird, wir bleiben bei HUXLEY, nur der „Wilde“, der noch die Emotionen von Shakespears Gesamtausgabe kennt – denn er wird etwas vermissen. Die „glückliche“ Mehrheit in der Matrix wird durch die digitalen Lösungen ihrer Wirkmächtigkeit auf die eigene Realität beraubt und kann ihren tieferen Lebenssinn nur mehr in simulierten Belustigungen finden. Ein Ausbrechen aus diesem System ist kaum möglich, da mit dem Versuch es zu verlassen *zunächst* auch die Fiktionen und Simulationen von Sinn verloren gehen. Diese Welt erscheint uns stabil und *vordergründig* wünschenswert; letztlich aber ist sie hohl.

Auswege aus dem Problem wurden schon oft gesucht. *Mündig ist nach ADORNO der, der für sich selbst spricht, weil er für sich selbst gedacht hat und nicht bloß nachredet.*⁵⁹ Die Vorschläge und Konsequenzen, die ADORNO daraus zieht, sind zwar nicht an heutige soziale Netze oder generative KI angepaßt, bleiben ihrem Geist nach aber gültig [Ado13], S. 228–229:

Ich könnte mir etwa denken, daß man auf den Oberstufen von höheren Schulen, aber wahrscheinlich auch von Volksschulen gemeinsam kommerzielle Filme besucht und den Schülern ganz einfach zeigt, welcher Schwindel da vorliegt, wie verlogen das ist; daß man in einem ähnlichen Sinn sie immunisiert gegen gewisse Morgenprogramme, wie sie immer noch im Radio existieren, in denen ihnen sonntags früh frohgemute Musik vorgespielt wird, als ob wir, wie man so schön sagt, in einer ›heilen Welt‹ leben würden, eine wahre Angstvorstellung im übrigen; oder daß man mit ihnen einmal eine Illustrierte liest und ihnen zeigt, wie dabei mit ihnen unter Ausnutzung ihrer eigenen Triebbedürftigkeit Schlitten gefahren wird.

Die Verwendung von ChatGPT als künstliche Intelligenz für die Beantwortung von Fragen, anstatt als ein wenig schlauere Suchmaschine, kann zu einer Lähmung der menschlichen Urteilskraft führen, weil diese durch fortgesetzte Nutzung des Hilfsmittels nicht mehr trainiert wird. Das erstaunt nicht. Will der digital arbeitende Mensch körperlich gesund bleiben, so muß er regelmäßig Sport treiben, während den Soldaten im altgriechischen Sparta noch seine Lebensweise fit hielt. Wir müssen fragen: Welche Hilfsmittel sind zweckmäßig für unser Überleben und welche führen zu einer solchen Entfernung von Lebensfähigkeit und Lebenssinn, daß wir diesen Verlust anschließend durch Surrogate wieder kompensieren müssen. Wenn wir kognitive Aufgaben ChatGPT überlassen – füllt dann neue Arbeitsverdichtung die gewonnene Zeit oder dient *binge watching* auf Netflix als Lückenfüller? Wenn wir einer KI die Routineaufgaben übertragen, woher beziehen wir die kognitive Fitness zur Bewältigung jener größeren Aufgaben, die wir dem System nicht, noch nicht oder grundsätzlich nicht übertragen wollen? Man mag an eine Karikatur des Menschen denken, der mit verkümmerten Muskeln im Rollstuhl sitzend von Robotern bedient wird – das einzige noch funktionierende Körperteil ist der Finger zur Betätigung von Joystick und Tastatur.

Anwendung auf die Wissenschaft Wissenschaft könne, so die problematische Annahme des Scientismus, das gesellschaftliche Leben steuern. Sie wird dabei weder durch Grundrechte noch durch Freiheitsrechte eingeschränkt, denn die scheinbar ultimative Ermächtigung durch die Wirklichkeit und ihre Zwänge entzieht sich schließlich demokratischer Meinungsbildung. Wollten wir etwa über Fakten oder Naturgesetze abstimmen? Werden Erkenntnisse nun von einem Algorithmus ermittelt oder von einer Maschine behauptet, so erscheinen sie auf besondere Weise objektiviert und legitimiert. Stehen deren Dienste nun allen zur Verfügung, so können sich alle selber, jederzeit und überall, über die „Wahrheit“ informieren. Zweifel

⁵⁹Hier zitiert nach [Hil23].

oder gar Kritik erhalten dadurch zwangsläufig den Status von „Desinformation“ und ihre Stigmatisierung als „Delegitimierung“ wirkt begründbar.

KI – wenn sie denn einmal die Kinderkrankheiten und Fehler überwunden haben wird – implementiert de facto einen normativen Standard mit absolutem Geltungsanspruch. Dieser ist gerade deshalb so gefährlich, weil generative KI eine Übernahme von Entscheidungen aus Gründen von Bequemlichkeit und Einfachheit bewirkt, ohne daß die Entscheider genügend Chance und Zeit für eine Reflexion haben, ob sie den dahinter verborgenen impliziten Ziel- und Wertesystemen überhaupt zustimmen.

Eine vorläufige Bilanz Wenn wir diesen Mechanismus nicht durchbrechen, so führt er von der Massenanzahl der künstlichen Intelligenz langfristig in eine totalitäre Gesellschaftsform. Aus der Rationalität der Aufklärung, die den Geist befreien wollte, wäre dann eine Diktatur durch den künstlichen Geist geworden. Wissenschaft sollte der Beschreibung und Modellierung von Beobachtungen dienen und nicht das psychologische Bedürfnis nach Sicherheit bedienen. KI kann ihr jedoch ein einfaches Hilfsmittel sein.

Nutzen wie auch Schaden einer künstlichen Intelligenz entstehen nur durch den Menschen und durch seine Bewertung ihres Einsatzes. Sicherlich ist sie ein Werkzeug, das manches leisten kann. Zwei Dinge aber kann künstliche Intelligenz nicht: *Verantwortung* übernehmen und *Sinn* finden. Beides wieder zu entdecken kann ein Ausweg werden.

Was wir von der Welt zu sehen bekommen bestimmt unsere Werte, unsere Werte legen fest, wie wir die unauflösbaren Paradoxa des Lebens betrachten wollen und es sind diese Entscheidungen, die letztlich unser Leben formen. Der Geist der Aufklärung kann uns zu einer eigenen Gestaltung dieser Aspekte aufrufen statt sie Ideologien, digitalen Portalen oder künstlichen Intelligenzen zu überlassen.

Mut für diesen Weg macht der Erfolg der historischen Aufklärung. Ihr Programm hat schon einmal funktioniert und es kann, entsprechend angepaßt, auch ein weiteres Mal Erfolg haben.

Eine mögliche Gegenposition Als Wissenschaftler sollten wir mögliche Gegenposition durchdenken und die für die Aufklärer höchst provokante Frage stellen: *Will* denn jeder selber denken und verstehen?

Unser wissenschaftliches Weltbild versucht, alles auf erste Prinzipien zu beziehen. Durch KI kann es hier zu einer gesellschaftlichen Verschiebung kommen: Möglicherweise werden in der nahen Zukunft immer weniger Menschen ein Verständnis des Universums anstreben um darauf ein besseres Überleben zu gründen. Wenn selbst die Erfinder der künstlichen Intelligenz ihre Produkte nicht mehr vollständig verstehen, dann wird die Akzeptanz des Nicht-Verstehens steigen. Wenn jeder gesellschaftlichen Schichte medial eine leicht dümmere Welt vorgespiegelt wird als jene, die sie selber gerade noch begreift, so kann jeder Mensch für sich die Illusion einer für ihn rationalen Weltsicht aufrecht erhalten und glauben, er verstünde wie die Welt funktioniert. Die Unterhaltungsindustrie mit ihren Reality Shows wie Big Brother oder Dschungelcamp ist bereits auf dem besten Weg in eine solche Gesellschaft – und für ihre Wissenschaftler könnte datengestützte KI die Rolle dieser Illusion übernehmen. HUXLEY erklärt die Stabilität solcher Gemeinschaften mit seinem Kasten-System, in dem Schlafschulen ihre Mitglieder konditionieren, mit ihrer Lebenslage zufrieden zu sein und sich als Individuen aufzugeben um sich in der anonymen Masse geborgen zu fühlen. Bienen und Ameisen bieten reale und funktionierende Beispiele solcher Gemeinschaften.

Die Apologeten rascher Veränderungen verursachen oft gesellschaftlichen Unfrieden. Sie können sich aber weder auf die angebliche Zwangsläufigkeit einer Entwicklung berufen noch auf die Unkenntnis der Gefahren ihres Tuns.

Literatur

- [Ado13] Theodor W. Adorno. *Erziehung zur Mündigkeit: Vorträge und Gespräche mit Hellmut Becker 1959-1969*. eBook Suhrkamp Verlag Berlin, 2013. ISBN: 978-3-518-73845-0.
- [AF21] Shruti Agarwal und Hany Farid. „Detecting Deep-Fake Videos from Aural and Oral Dynamics“. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2021, S. 981–989. DOI: [10.1109/CVPRW53098.2021.00109](https://doi.org/10.1109/CVPRW53098.2021.00109).
- [Bri18] James Bridle. *New Dark Age: Technology and the end of the Future*. Verso, 2018. ISBN: 978-1-78663-547-1.
- [Bro91] Rodney A. Brooks. „Intelligence without Reason“. In: *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*. 1991, S. 569–595.
- [Cap11] Clemens H. Cap. „Content Neutrality for Wiki Systems: From Neutral Point of View (NPOV) to Every Point of View (EPOV)“. In: *Proceedings of the Fourth International Conference on Internet Technologies and Applications (ITA 11), Wrexham, UK, 6.- 9. September 2011*. 2011. URL: https://www.researchgate.net/publication/288629410_Content_neutrality_for_Wiki_systems_From_neutral_point_of_view_NPOV_to_every_point_of_view_EPOV.
- [Cap12] Clemens H. Cap. „Towards Content Neutrality in Wiki Systems“. In: *Future Internet 4.4* (Dez. 2012), S. 1086–1104. ISSN: 1999-5903. DOI: [10.3390/fi4041086](https://doi.org/10.3390/fi4041086).
- [Cap17] Clemens H. Cap. „Vertrauen in der Krise: Vom Feudalismus 2.0 zur Digitalen Aufklärung“. In: *Öffentliches Vertrauen in der Mediengesellschaft*. Hrsg. von Michael Haller. Halem-Verlag, 2017, S. 237–254. ISBN: 978-3-86962-099-2.
- [Cap19] Clemens H. Cap. „Feudalismus oder Aufklärung? Optionen der digitalen Gesellschaft“. In: *Die Maschine: Freund oder Feind?* Hrsg. von Caja Timm und Thomas Christian Bächle. Springer Fachmedien Wiesbaden, 2019, S. 161–182. ISBN: 9783658229542. DOI: [10.1007/978-3-658-22954-2_8](https://doi.org/10.1007/978-3-658-22954-2_8).
- [Cap23] Clemens H. Cap. „„Der neue Gott ist nackt!“. ChatGPT im Bildungswesen“. In: *Forschung und Lehre 5* (2023), S. 344–345. ISSN: 0945-5604. DOI: [10.37307/j.0945-5604.2023.05.09](https://doi.org/10.37307/j.0945-5604.2023.05.09). URL: <https://www.forschung-und-lehre.de/zeitfragen/der-neue-gott-ist-nackt-5604>.
- [DB20] Louisa Dahmani und Véronique D. Bohbot. „Habitual use of GPS negatively impacts spatial memory during self-guided navigation“. In: *Scientific Reports 10.1* (Apr. 2020). ISSN: 2045-2322. DOI: [10.1038/s41598-020-62877-0](https://doi.org/10.1038/s41598-020-62877-0).
- [GVA21] Nadia Guerouaou, Guillaume Vaiva und Jean-Julien Aucouturier. „The shallow of your smile: The ethics of expressive vocal deep-fakes“. In: *Philosophical Transactions of the Royal Society B: Biological Sciences 377*.1841 (Nov. 2021). DOI: <https://doi.org/10.1098/rstb.2021.0083>. URL: <https://doi.org/10.1098/rstb.2021.0083>.
- [HA20] Max Horkheimer und Theodor W. Adorno. *Dialektik der Aufklärung: Philosophische Fragmente*. Fischer-TB, 2020. ISBN: 3596274044.
- [Han14] Byung-Chul Han. *Müdigkeitsgesellschaft*. 10. Auflage. Matthes & Seitz Berlin, 2014. ISBN: 978-388221-616-5.
- [HC88] Edward S. Herman und Noam Chomsky. *Manufacturing Consensus: The Political Economy of the Mass Media*. Pantheon, 1988. ISBN: 0375-71449-9.
- [Hil23] Veronika Hilzensauer. „Widerständiges Denken als Ausdruck von Mündigkeit. Gedanken mit Theodor W. Adorno zum Zeitalter des Anthropozäns“. In: (März 2023), S. 98–101. URL: https://kath-akademie-bayern.de/wp-content/uploads/Hilzensauer_Widerstand_2022.pdf.
- [Lew79] Clive Staples Lewis. *God in the Dock*. Collins, 1979. ISBN: 0006253717.
- [Lie17] Konrad Liessmann. *Bildung als Provokation*. Zsolnay Verlag, 2017. ISBN: 978-3-552-05824-8.
- [Lüb89] Hermann Lübbe. *Politischer Moralismus: Der Triumph der Gesinnung über die Urteilskraft*. Siedler, 1989.

⁶⁰Hinweis: Die Internet-Quellen beziehen sich auf den 15. März 2024 als Abrufdatum.

- [MW07] Anil Markandya und Paul Wilkinson. „Electricity generation and health“. In: *The Lancet* 370.9591 (2007), S. 979–990. DOI: [10.1016/S0140-6736\(07\)61253-7](https://doi.org/10.1016/S0140-6736(07)61253-7).
- [OM54] James Olds und Peter Milner. „Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain.“ In: *Journal of Comparative and Physiological Psychology* 47.6 (1954), S. 419–427. ISSN: 0021-9940. DOI: [10.1037/h0058775](https://doi.org/10.1037/h0058775).
- [Ott09] Max Otte. *Der Informationscrash*. 1. Auflage. Econ, 2009. ISBN: 3430200784.
- [PB06] Rolf Pfeifer und Josh Bongard. *How the Body Shapes the Way We Think*. MIT Press, 2006. ISBN: 9780262537421. DOI: [10.7551/mitpress/3585.001.0001](https://doi.org/10.7551/mitpress/3585.001.0001).
- [Pop89] Karl Raimund Popper. *Logik der Forschung*. 9. Auflage. J. C. B. Mohr Tübingen, 1989. ISBN: 3-16-345484-4.
- [Pop92] Karl Raimund Popper. *Die offene Gesellschaft und ihre Feinde. Band 2*. Mohr / Siebeck, Tübingen, 1992. ISBN: 9783161459528.
- [Put69] Hilary Putnam. „Is logic empirical?“ In: *Boston Studies in the Philosophy of Science: Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*. Springer, 1969, S. 216–241.
- [Roz23] David Rozado. „The Political Biases of ChatGPT“. In: *Social Sciences* 12.3 (2023). ISSN: 2076-0760. DOI: [10.3390/socsci12030148](https://doi.org/10.3390/socsci12030148). URL: <https://www.mdpi.com/2076-0760/12/3/148>.
- [SL09] Michael Schmidt und Hod Lipson. „Distilling Free-Form Natural Laws from Experimental Data“. In: *Science* 324.5923 (2009), S. 81–85. DOI: [10.1126/science.1165893](https://doi.org/10.1126/science.1165893).
- [Sov+16] Benjamin K. Sovacool u. a. „Balancing safety with sustainability: assessing the risk of accidents for modern low-carbon energy systems“. In: *Journal of Cleaner Production* 112 (Jan. 2016), S. 3952–3965. ISSN: 0959-6526. DOI: [10.1016/j.jclepro.2015.07.059](https://doi.org/10.1016/j.jclepro.2015.07.059).
- [Ste16] Anna Steiner. „Künstliche Intelligenz: Zum Nazi und Sexisten in 24 Stunden“. In: *Frankfurter allgemeine Zeitung* (24. 03. 2016). URL: <https://www.faz.net/aktuell/wirtschaft/netzwirtschaft/microsofts-bot-tay-wird-durchnutzer-zum-nazi-und-sexist-14144019.html>.
- [Ste69] Wolfgang Stegmüller. *Metaphysik, Skepsis, Wissenschaft*. 2. Auflage. Springer-Verlag Berlin, 1969.
- [Vol92] Gerhard Vollmer. „Die vierte bis siebte Kränkung des Menschen? Gehirn, Evolution und Menschenbild“. In: *Philosophia Naturalis* 29.1 (1992), S. 118–134.
- [Wac13] Thomas Wachtendorf. „Aber der Löwe spricht eben nicht! Anmerkungen zu einer Kontroverse“. In: *The Wittgenstein Archives at the University of Bergen (WAB)* (2013). URL: <https://wab.uib.no/agora/tools/alws/collection-6-issue-1-article-8.annotate>.
- [Wei78] Joseph Weizenbaum. *Die Macht der Computer und die Ohnmacht der Vernunft*. Suhrkamp taschenbuch wissenschaft, 1978. ISBN: 3-518-27874-6.
- [Wie79] Anna Wierzbicka. „Ethno-syntax and the philosophy of grammar“. In: *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 3.3 (1979), S. 313–383.
- [Wit03] Ludwig Wittgenstein. *Philosophische Untersuchungen*. Suhrkamp, 2003. ISBN: 978-3-518-22372-7.
- [WWF75] Paul Watzlawick, John H. Weakland und Richard Fisch. *Lösungen: Zur Theorie und Praxis menschlichen Wandels*. 3. unveränd. Edition. Hogrefe AG, 1975. ISBN: 978-3456800387.
- [Yan+24] Yuchen Yang u. a. „SneakyPrompt: Jailbreaking Text-to-image Generative Models“. In: *2024 IEEE Symposium on Security and Privacy (SP)*. Als Preprint zugänglich unter <https://arxiv.org/abs/2305.12082>. Los Alamitos, CA, USA: IEEE Computer Society, Mai 2024, S. 122–122. DOI: [10.1109/SP54263.2024.00123](https://doi.org/10.1109/SP54263.2024.00123). arXiv: [2305.12082 \[cs.LG\]](https://arxiv.org/abs/2305.12082). URL: <https://doi.ieeecomputersociety.org/10.1109/SP54263.2024.00123>.