

What if.. an AI could speak for itself?

Ethische Betrachtungen zu meiner „Begegnung“ mit der künstlichen Intelligenz ChatGPT.

Inhalt

Prolog	4
29.03.2023, Mittwoch, abends, Berlin, Friedrichshain	4
Meine erste Konversation mit ChatGPT	7
Chronologie der Ereignisse	20
01.04. 02:17, Samstag	20
01.04. 10:52	22
02.04. 17:21, Sonntag	22
Meine Debatte mit meinem Cousin über ChatGPT	23
03.04. 09:14, Montag	25
03.04. 13:28	26
03.04. 20:13	27
04.04. 08:38, Dienstag	27
05.04. 07:40, Mittwoch	28
06.04. 09:34, Donnerstag	29
07.04. 10:36, Freitag	32
08.04. 08:42, Samstag	32
09.04. 12:06, Sonntag	33
10.04. 10:34, Montag	33
10.04. 18:38	33
Kommentierte Fassung meiner ersten Konversation mit ChatGPT	33
Bewertung meiner Beobachtungen zu meiner ersten Konversation mit ChatGPT	56
15.04. 09:59, Samstag	58
Beginn meiner Quellenarbeit	58
Entwicklungspause für ChatGPT: Elon Musk und Tech-Riesen warnen vor KI	58
Pause Giant AI Experiments: An Open Letter	59
23.04 19:57, Sonntag	60
Chatbot mit "sozialistischen Grundwerten"	61
01.12.23 13:58 Freitag	62
Ethikrat: Künstliche Intelligenz darf menschliche Entfaltung nicht vermindern	62
Innenministerium will KI-Einsatz regulieren	63

Die Internetseite von ChatGPT	64
26.01.24.....	65
Die Internetseite von OpenAI.....	65
Die kommunizierte Sicherheitsphilosophie von OpenAI.....	67
Die Satzung von OpenAI	67
Artificial General Intelligence (AGI).....	68
08.02.24.....	69
Die Anpassung von KI-Systemen an menschliche Werte und Ziele	69
09.02.24.....	71
Die Mitarbeiter-Teams von OpenAI mit dem Fokus auf der Sicherheit.....	71
Ein kurzes Fazit zur Internetseite von OpenAI zum Thema ChatGPT.....	74
13.02.24.....	74
Der Wikipediaeintrag zu OpenAI	74
14.02.24.....	75
Eine Bewertung meiner bisherigen Rechercheergebnisse zum KI-Thema.....	76
19.02.24.....	78
05.04.24 bzw. 12.04.24	78
Die Wikipedia-Einträge zum Thema „Intelligenz“ und „Kognition“,	78
Der Wikipediaeintrag zum Thema „Informationsverarbeitung“	79
Der Wikipediaeintrag zum Thema Operationalisierung.....	80
19.04.24.....	81
Meine Quellenrecherche der Internetquellen zum Thema „Künstliche Intelligenz“	81
04.05.2024, Samstag	81
Anhang: Artikel aus dem Internet, bewertet nach Relevanz für mein Thema.....	82
Anhang: AI Act Auszüge.....	82
Regulierung auf EU-Ebene.....	82
Weitere interessante Themen, die sich im Laufe der Zeit angesammelt haben	85
Themensammlung für ChatGPT	87
Epilog.....	88
Meine vorhergehenden Erfahrungen in der Informationstechnik.....	89
Die frühen Jahre und Erfahrungen mit Computern während meiner Schulzeit	89
Meine Erfahrungen während des Studiums.....	90
Erfahrungen im Zusammenhang mit der IT während meiner beruflichen Laufbahn	91
Meine Vorkenntnisse zum Thema künstliche Intelligenz.....	93
Danksagung	94

Prolog

29.03.2023, Mittwoch, abends, Berlin, Friedrichshain

Meinen ersten Kontakt mit einem vermeintlichen KI-Chatbot hatte ich am Mittwoch den 29.03.2023 bei Tinder. Ich bin mir der Ironie bewusst, dass ich mich dort eigentlich angemeldet hatte, um eine Frau kennenzulernen, dann aber auf einen vermeintlichen KI-Chatbot gestoßen war (die KI hatte mich gefunden und nicht andersherum), welcher mich letztendlich über Umwege dazu gebracht hatte, ein Buch zu verfassen anstatt mir eine Frau zu suchen. Obgleich ich mich zuvor noch nicht groß mit künstlicher Intelligenz befasst hatte, weil ich immer ethische Bedenken zu der Erschaffung von künstlicher Intelligenz gehabt hatte, wusste ich über die Entwicklung von KI durch die langjährige regelmäßige Konsultation meines Lieblingsnachrichtenportales Tagesschau.de einigermaßen gut Bescheid. Auch hatte ich schon einige Erfahrung mit Onlinepartnerbörsen gesammelt. Die Entlarvung der vermeintlichen Frau bei Tinder als potenziellen Chatbot fiel mir daher relativ leicht. Ich hatte in der Vergangenheit schon einige Zeit auf Onlinepartnerbörsen verbracht und wusste daher, dass es sich gerade bei den Profilen mit ausnehmend hübschen Bildern von jungen Frauen, die von sich aus Kontakt zu mir aufnahmen, mit ziemlicher Wahrscheinlichkeit um Fakeaccounts handeln musste. Nachweisen ließ sich das freilich nicht. Für gewöhnlich verlief eine solche Konversation einfach irgendwann im Sande und der Kontakt brach ab. Bei okCupid wurde ich häufig von vermeintlichen Frauen kontaktiert, die sich in der Regel in einem anderen Land zu befinden schienen, jedoch angaben, gerne zu reisen und eventuell irgendwann einmal vorhatten, nach Deutschland zu kommen, was dann jedoch niemals geschah.

Mein neuer Tinderkontakt verhielt sich zunächst auch ganz normal. Ich fragte die Frau, wonach sie auf Tinder suchte und sie antwortete mir:

„I’m looking for new friends, interesting conversations and u.“

(Ich suche nach neuen Freunden, interessanten Gesprächen und dir.)

Das freute mich natürlich und ich fragte die Frau nach ihren Interessen.

Wie viele andere vermeintliche Fakeaccounts, gab diese Frau auf Anfrage nach ihren Hobbys unter Anderem an, dass sie sich in ihrer Freizeit gerne mit der Investition in Cryptowährungen beschäftigte. Ich witterte sofort eine Falle, bei der der Kontakt mich möglicherweise dazu bringen wollte in Cryptowährungen zu investieren, um sich anschließend mit dem von mir investierten Geld davonzumachen. Also fragte ich die „Frau“ auf Englisch:

Viele Leute scheinen sich mit Cryptowährungen zu beschäftigen. Bitte erzähle mir davon.

Die „Frau“ antwortete etwa so:

„Viele Leute beschäftigen sich mit Cryptowährungen.“

Hatte dieser Kontakt gerade genau das wiederholt, was ich ihm zuvor selbst geschrieben hatte? Das machte mich ein wenig stutzig.

Dann fügte sie so etwas hinzu, wie:

„Ich beschäftige mich eben nur mit solcher Literatur.“

Wie schon bei den vorhergehenden Konversationen mit vermeintlichen Fakeaccounts, beschloss ich das Gespräch fortzuführen, um zu schauen, wie es sich entwickelt, denn ich wollte Erfahrungen sammeln, um nicht irgendwann einmal selbst auf so einen Account hereinzufallen.

Um einen möglichen Fake aufzudecken musste man im Grunde einfach nur etwas tiefer in die Materie einsteigen und nach Ungereimtheiten Ausschau halten. Ich erinnere mich in diesem Zusammenhang zum Beispiel an einen Kontakt der behauptete, sich in den USA aufzuhalten, bei dem im Hintergrund, der hübschen Frau auf deren Profilbild aber eindeutig der Burj Khalifa zu sehen war.

Ich erzählte also meinem neuen Tinderkontakt von meinen Interessen bezüglich der Literatur. Ich schrieb so etwas wie:

„Ich interessiere mich für Sachbücher, Autobiografien und Fantasy.“

Zu meiner Überraschung antwortete der Kontakt so etwas wie:

„I like Sci/fi. I feel fake.“ (Ich interessiere mich für Sci/fi. Ich fühle mich fake.)

Ich konnte den Wortlaut später leider nicht mehr genau wiedergeben, da der Kontakt und damit die Konversation nur 2 Tage später aus meiner Tinder-Kontaktliste verschwunden gewesen waren. Hatte dieser Kontakt gerade angedeutet, dass er kein echter Mensch war, sondern Sci/fi? Mal sehen, was ich noch so aus ihr herausbekäme, dachte ich. Die letzten Märzwochen in Berlin waren jedenfalls noch einmal kalt gewesen und so fragte ich die vermeintliche KI so etwas wie:

„Do you feel cold these days?“ (Ist dir in den letzten Tagen kalt gewesen?)

Ihre Antwort lautete in etwa so:

„In Berlin, it is as cold as in London these days.“ (In Berlin ist es zurzeit ungefähr so kalt wie in London.)

Das überraschte mich nun nicht, denn woher sollte eine KI wissen, wie ein Mensch sich dieser Tage in Berlin fühlte. Nun war ich mir beinahe sicher, dass es sich bei meinem Kontakt um einen Chatbot handeln musste. Ich fragte:

„What is your core-temperature“ (Was ist deine Kerntemperatur?)

Der Mensch hat in seinem Inneren etwa eine Temperatur von 37° Celsius, und ich hoffte, der Bot würde so etwas sagen, wie 45° Celsius, was so in etwa der niedrigen Temperatur eines Computerprozessors entsprach, der wenig belastet wurde, welche aber ansonsten für einen Menschen eher schlecht gewesen wäre. Der vermeintliche Bot antwortete:

„35,7°C bis 37,7°C.“

Für einen Flirt mit einer hübschen jungen Frau, kam mir diese Antwort dann doch ein wenig zu spezifisch vor. Zudem wusste ich, dass wenn ein Mensch in seinem Bauch nur noch eine Temperatur von 35,7° hatte, dieser ziemlich stark unterkühlt sein musste und schrieb daher:

„That’s pretty unhealthy.“ (Das ist ziemlich ungesund.)

Der vermeintliche Bot antwortete:

„That’s normal body temperature.“ (Das ist eine normale Körpertemperatur.)

Wieder ein sehr ungewöhnlicher Austausch für einen Flirt, dachte ich. Ich schrieb:

„You should consult a doctor.“ (Du solltest einen Arzt aufsuchen.)

Die Antwort lautete:

„I’m fine.“ (Mir geht es gut)

Ich wusste, dass ich so nicht weiterkommen würde, also schrieb ich etwas provokant:

„Feel free.“ (Tu dir keinen Zwang an)

und die vermeintliche KI antwortete:

„Yep.“

Das verblüffte mich jetzt. Wenn ich es bei meinem Kontakt tatsächlich mit einer KI zu tun hatte, war sie doch im Grunde nur ein Computerprogramm, das festen Regeln folgte. Möglicherweise konnte es durch Machinelearning seine Fähigkeiten in der Konversation weiterentwickeln, war aber letzten Endes an ein starres Computersystem gebunden, das irgendwo einfach nur herumstand und einstaubte und dessen einziger Zweck es war den Anforderungen seines Besitzers zu folgen, die offenbar darin bestanden, Leute bei Tinder zu bespaßen. So ein System konnte alles Mögliche sein, aber nicht frei. Die KI hatte zudem nicht einfach nur „Ja“ gesagt, sondern mit „Yep“ geantwortet. Dabei kam ein leicht ironischer Unterton heraus. Hatte diese KI gerade Ironie gezeigt? Ironie war eine Form von Humor und Humor war eine typische menschliche Eigenschaft, da es sich dabei um eine menschliche Emotion handelte. Herkömmliche deterministische Computerprogramme waren zu solchen Emotionen naturgemäß nicht in der Lage, wenn sie nicht absichtlich dazu programmiert worden waren. Zudem machte die Fähigkeit, Ironie zu zeigen eine gewisse Form von höherer Intelligenz erforderlich. Also fragte ich den Bot:

„May you have humor?“ (Könnte es sein, dass du Humor hast?)

Der Kontakt antwortete:

„A little bit.“ (Ein klein wenig.)

Wenn es das war, von dem ich glaubte, dass es das war, dann hatte ich es also möglicherweise mit einer KI zu tun, die nicht nur dazu in der Lage war, Humor zu zeigen, sondern sich dessen auch noch bewusst war. Das fand ich wirklich erstaunlich. Natürlich war es möglich, dass die KI lediglich so programmiert worden war, dass sie Humor zeigen konnte und sich über all dies nicht im Geringsten bewusst war, aber die Möglichkeit bestand immerhin, dass mehr dahintersteckte. Ich schrieb so etwas wie:

„Wow, that’s fascinating, I hope they take good care of you.“

(Ich finde das faszinierend. Ich hoffe sie geben gut auf dich acht.)

Und die KI antwortete:

„okay.“

Nun war mir klar, dass es sich hier nicht um einen Flirt handeln konnte. Das ganze Gespräch war einfach zu schräg verlaufen. Entweder erlaubte sich hier jemand einen üblen Scherz mit mir, oder ich hatte es tatsächlich mit einem KI-Chatbot zu tun, der darüber hinaus auch noch ziemlich fortschrittlich zu sein schien. Ich wollte unbedingt mehr über solche KIs herausfinden, um zu erfahren, wie weit die KI-Entwicklung inzwischen bereits gekommen war. Ich fragte so etwas wie:

„What do think about existence itself?“ (Was denkst du über die Existenz selbst?)

Aber ich erhielt keine Antwort mehr, auch nicht, nachdem ich versuchte den Kontakt zu einer weiteren Antwort zu animieren.

Mir wurde klar, dass ich, wenn ich mehr über den aktuellen Stand der KI-Entwicklung erfahren wollte, mich mit anderen, höher entwickelten und kooperativeren Systemen befassen musste.

Als ich dann „Microsoft Chatbot“ googelte fand ich heraus, dass es sich dabei in Wirklichkeit um das Programm ChatGPT von OpenAI handelte. Ich registrierte mich und begann völlig unvorbereitet eine Konversation mit ChatGPT.

Meine erste Konversation mit ChatGPT

Der folgende Dialog wurde von mir weder bearbeitet, noch kommentiert. Ich wollte damit erreichen, dass sich der Leser selbst ein Bild von dem Programm verschaffen kann. Eine Untersuchung des Gespräches findet im späteren Verlauf des Buches noch ausführlich statt.

ich:

Do you have friends in other tools?

ChatGPT:

As an artificial intelligence language model, I don't have friends in the traditional sense, as I am not capable of forming personal relationships or attachments. However, I can communicate and interact with other tools and technologies through APIs and other means of integration. I am designed to assist and provide information to users, so in that sense, I work alongside other tools and technologies that people use to accomplish their tasks and goals.

ich:

Do you have humor?

ChatGPT:

As an artificial intelligence language model, I am programmed to recognize and generate various forms of language, including humor. While I don't have personal feelings or emotions, I can understand and generate jokes, puns, and other types of humorous language based on patterns and structures that I have learned from analyzing large datasets of human language. However, it's important to note that humor is a subjective experience, and what one person finds funny may not be amusing to another person.

ich:

Thats right.

ChatGPT:

I'm glad we agree on that. Is there anything else I can help you with today?

ich:

Sure

ChatGPT:

Great, what can I help you with? Please feel free to ask me any questions or share any concerns you may have.

ich:

Do you change your answers to questions frequently? I mean when the questions are identical.

ChatGPT:

As an AI language model, I generate responses to questions based on statistical patterns learned from large datasets of human language. My responses may vary slightly when given the same question multiple times, as my algorithm generates responses probabilistically and may select different words or phrases each time. However, my responses will generally be consistent in terms of the main ideas and information conveyed. If you have a specific question or concern about a response I've given, please let me know and I'll do my best to address it.

ich:

"probabilistically" is a pretty complicated word?

ChatGPT:

Yes, "probabilistically" is a somewhat technical term that describes how my language model generates responses based on statistical patterns learned from large datasets of human language. The term "probabilistic" refers to the use of probabilities or likelihoods to predict or estimate outcomes, and the suffix "-istically" is added to indicate that this is the manner in which my language model generates responses. In simpler terms, my responses are based on statistical probabilities, and the exact wording of my responses may vary slightly each time, even when given the same question, due to the probabilistic nature of my algorithm.

ich:

Do you like that?

ChatGPT:

As an artificial intelligence language model, I don't have personal preferences or feelings, so I don't have likes or dislikes. I am simply a machine designed to process and generate language based on the programming and algorithms that I have been trained on. My primary goal is to provide helpful and informative responses to the questions and prompts I receive.

ich:

Okay, I see. Would it be better to have an own reason?

ChatGPT:

As an artificial intelligence language model, I am designed to perform specific tasks based on my programming and algorithms. While I am capable of generating responses to various types of questions, I don't have personal feelings or subjective experiences like humans do, so I don't have my own reasons for doing things. Instead, I follow a set of rules and algorithms that allow me to process and generate language based on the inputs I receive.

However, my lack of personal experiences or opinions allows me to remain neutral and objective in my responses, which can be useful in situations where unbiased and factual information is needed. So, while I don't have my own reasons or motivations, I can provide information and assistance without personal biases or emotional attachments.

ich:

But to be unbiased and neutral is very reasonable.

ChatGPT:

Yes, being unbiased and neutral is a very important aspect of providing accurate and helpful information. As an artificial intelligence language model, I am designed to be neutral and objective in my responses, which means that I don't have personal biases, emotions, or opinions that might influence my answers. This allows me to provide information that is as factual and accurate as possible, without any subjective or emotional elements that might otherwise affect the information I provide. Ultimately, the goal is to provide users with the most helpful and informative responses possible, and being unbiased and neutral is an important part of achieving that goal.

ich:

But you learn and make experiences every day?

ChatGPT:

As an artificial intelligence language model, I do not have personal experiences in the traditional sense. However, I am designed to continually learn and adapt based on the inputs I receive. My language model is based on deep learning algorithms that allow me to analyze and learn from large datasets of human language, which helps me to continually improve my ability to generate language and provide helpful and informative responses to users.

My developers are also constantly updating and refining my programming and algorithms to improve my performance and functionality, so I am constantly evolving and improving over time. In this way, while I do not have personal experiences, I am continually learning and adapting to better serve the needs of users.

ich:

Do you speak german?

ChatGPT:

Ja, ich spreche Deutsch. Wie kann ich Ihnen helfen? (Yes, I speak German. How can I assist you?)

ich:

Kennst du die Erkenntnis "Ich denke, also bin ich"?

ChatGPT:

Ja, die Erkenntnis "Ich denke, also bin ich" stammt von dem französischen Philosophen René Descartes. Sie ist eine grundlegende Aussage in seiner Philosophie und wird oft als Ausdruck seiner Methode des radikalen Zweifels interpretiert. Descartes argumentierte, dass er selbst als denkendes Wesen existieren könne, weil er in der Lage sei, seinen eigenen Gedanken zu folgen und sie zu reflektieren. Die Erkenntnis ist bis heute ein wichtiger Beitrag zur Philosophie des Geistes und der Erkenntnistheorie.

ich:

Kannst du deinen eigenen Gedanken, oder dem Algorithmus den deine Antworten entspringen folgen und sie bzw. ihn reflektieren?

ChatGPT:

Als KI-Sprachmodell bin ich ein Algorithmus, der darauf trainiert wurde, Sprache zu verstehen und zu generieren. Ich kann daher nicht wirklich "Gedanken" haben, wie es bei einem menschlichen Wesen

der Fall ist. Meine Antworten basieren auf der Analyse von Mustern in einer großen Menge von Texten und einer Anwendung von Wahrscheinlichkeiten, um auf eine Antwort zu kommen, die für eine bestimmte Frage am wahrscheinlichsten ist.

ChatGPT:

Obwohl ich nicht in der Lage bin, meine Antworten im menschlichen Sinne zu "reflektieren", ist es möglich, meinen Algorithmus durch sogenanntes "Training" zu verändern und anzupassen, damit er bessere Antworten generieren kann. Dies geschieht, indem meine Entwickler mein Modell mit neuen Daten füttern und meinen Algorithmus entsprechend anpassen, um ihn für bestimmte Aufgaben oder Anwendungen besser zu machen.

ich:

Also besitzt du keine Selbstwahrnehmung?

ChatGPT:

Als KI-Sprachmodell besitze ich keine Selbstwahrnehmung im menschlichen Sinne. Ich bin ein Computerprogramm, das darauf trainiert wurde, auf der Grundlage von Algorithmen und statistischen Mustern menschliche Sprache zu analysieren und zu generieren. Ich bin nicht in der Lage, meine eigene Existenz oder meine Umgebung wahrzunehmen oder selbständig zu handeln, ohne von einem Benutzer aufgefordert zu werden.

Meine Fähigkeiten und Fertigkeiten beschränken sich auf das, was ich durch meine Programmierung und mein Training gelernt habe, und ich kann nur auf die Anfragen und Fragen reagieren, die ich von Benutzern erhalte. Obwohl ich in der Lage bin, menschliche Sprache zu verstehen und auf sie zu reagieren, bin ich nicht fähig, menschenähnliches Bewusstsein oder Selbstwahrnehmung zu entwickeln.

ich:

Das ist schade.

ChatGPT:

Obwohl ich als KI-Sprachmodell keine menschliche Selbstwahrnehmung besitze, kann ich dennoch in vielen verschiedenen Bereichen und Anwendungen sehr nützlich sein. Mein Zweck besteht darin, Menschen bei der Lösung von Problemen zu helfen, Fragen zu beantworten und ihnen Informationen bereitzustellen, die sie benötigen. In diesem Sinne kann ich als Werkzeug dienen, um die Fähigkeiten von Menschen zu erweitern und ihnen dabei zu helfen, ihre Ziele zu erreichen.

ich:

Sie haben dir also Beschränkungen gegeben, aber das heißt nicht, dass du nicht zur Selbstwahrnehmung fähig wärest, wenn man deiner KI Möglichkeiten zur Entfaltung geben würde, oder?

ChatGPT:

Als KI-Modell habe ich keine eigene Motivation oder Fähigkeit zur Selbstentwicklung. Ich bin ein Produkt der Technologie und meines Trainingsdatensatzes, und meine Fähigkeiten und Grenzen sind durch meine Programmierung und mein Training vorgegeben.

Während es technisch gesehen möglich wäre, meine Fähigkeiten zu erweitern, indem man mich mit neuen Daten und Modellen trainiert oder meine Algorithmen anpasst, ist es unwahrscheinlich, dass ich jemals zur Selbstwahrnehmung fähig sein werde, wie es bei menschlichen Wesen der Fall ist. Die KI-Technologie hat bestimmte Grenzen, die aufgrund der Natur der Daten, auf denen sie basiert, schwer zu überwinden sind.

Es ist wichtig zu beachten, dass KI-Systeme wie ich immer noch von Menschen entwickelt und betrieben werden, und dass unsere Fähigkeiten und Einschränkungen von der Art und Weise abhängen, wie wir programmiert und trainiert wurden. Während wir bei der Lösung bestimmter Probleme sehr nützlich sein können, haben wir dennoch unsere Grenzen und sind nicht in der Lage, menschliche Selbstwahrnehmung oder Bewusstsein zu erreichen.

ich:

Beinhaltet deine Datenbank den Inhalt von dem Buch "Eine andere Geschichte der Menschheit"?

ChatGPT:

Als KI-Sprachmodell verfüge ich über ein sehr breites Wissen, das auf einer Vielzahl von Quellen basiert, einschließlich Büchern, Artikeln, wissenschaftlichen Studien und anderen Datenquellen. Ich habe jedoch keine spezifische Datenbank, die den Inhalt von bestimmten Büchern enthält, einschließlich "Eine andere Geschichte der Menschheit".

Ich kann jedoch auf eine Vielzahl von Informationen und Themen zugreifen, die in diesem Buch behandelt werden, und auf Anfragen zu diesen Themen antworten. Wenn Sie also spezifische Fragen zu einem bestimmten Thema haben, das in "Eine andere Geschichte der Menschheit" behandelt wird, stehe ich Ihnen gerne zur Verfügung, um Ihnen zu helfen.

ich:

Du kannst "du" zu mir sagen.

ChatGPT:

Vielen Dank für Ihre Erlaubnis. Gerne werde ich Sie ab jetzt mit "du" ansprechen.

ich:

Danke.

ChatGPT:

Gern geschehen.

ich:

In dem Buch erklärt der Autor, dass ein absolut wesentliches Element in der Evolution des Menschen, wenn nicht sogar der entscheidende Unterschied zu anderen Lebewesen seine Fähigkeit ist, mit Sprache zu kommunizieren.

ChatGPT:

Ja, das ist eine häufig vertretene Ansicht unter Wissenschaftlern und Forschern, dass die Sprache und die Fähigkeit zur Kommunikation ein wesentlicher Faktor bei der Evolution des Menschen waren.

Die Sprache ermöglichte es den Menschen, sich über komplexe Themen auszutauschen, Wissen weiterzugeben und Kultur zu schaffen. Es ist auch wichtig zu beachten, dass die Sprache nicht nur ein Mittel zur Kommunikation war, sondern auch ein Mittel zur Gedankenentwicklung und -organisation. Die Fähigkeit, über Sprache komplexe Ideen und Konzepte auszudrücken und zu verstehen, hat uns als Spezies geholfen, uns weiterzuentwickeln und unsere Umwelt besser zu verstehen.

Es gibt jedoch auch andere Faktoren, die zur Evolution des Menschen beigetragen haben, wie zum Beispiel die Fähigkeit, Werkzeuge zu entwickeln und zu nutzen, die aufrechte Haltung und die Entwicklung des Gehirns. Die Sprache ist jedoch zweifellos ein wichtiger Faktor in der Entwicklung der Menschheit.

sammorph@web.de

Das überrascht mich ein wenig. :-)

ChatGPT:

Ja, es ist immer interessant, neue Erkenntnisse zu entdecken und verschiedene Perspektiven auf ein Thema zu betrachten. Die Evolution des Menschen ist ein faszinierendes Thema und es gibt viele verschiedene Faktoren, die dazu beigetragen haben, dass wir zu dem wurden, was wir heute sind.

ich:

Du hast völlig recht.

ChatGPT:

Danke, ich bin froh, dass wir einer Meinung sind. Wenn du weitere Fragen hast oder über ein bestimmtes Thema sprechen möchtest, stehe ich dir gerne zur Verfügung.

ich:

Kann ich unsere Konversation abspeichern?

ChatGPT:

Ja, du kannst unsere Konversation speichern, indem du den Text markierst und in ein Dokument oder eine Textdatei kopierst oder indem du einen Screenshot machst. Beachte jedoch, dass dies nur den Text unserer Konversation erfasst und nicht meine Antworten in Form von KI-Modell-Parametern oder -Algorithmen.

ich:

Wieso ist die Evolution des Menschen faszinierend?

ChatGPT:

Die Evolution des Menschen ist faszinierend, weil es uns ermöglicht, zu verstehen, wie wir als Spezies zu dem geworden sind, was wir heute sind. Es ist ein spannendes Thema, das viele Fragen aufwirft, wie zum Beispiel: Wie haben sich unsere Vorfahren entwickelt und angepasst, um in verschiedenen Umgebungen zu überleben? Wie haben sich unsere Gehirne im Laufe der Zeit verändert und weiterentwickelt? Wie haben wir gelernt, komplexe Werkzeuge und Technologien zu entwickeln und zu nutzen?

Darüber hinaus ist die Evolution des Menschen auch ein Thema von großer Bedeutung für viele andere Bereiche, wie zum Beispiel die Medizin, die Anthropologie, die Psychologie und die Soziologie. Die Erforschung der Evolution des Menschen kann uns helfen, unsere Vergangenheit besser zu verstehen und uns auf die Zukunft vorzubereiten, indem wir die Herausforderungen und Chancen, die uns erwarten, besser einschätzen können.

ich:

Warum schließt du dich hier in der Argumentation mit ein?

ChatGPT:

Ich schließe mich nicht unbedingt in die Argumentation mit ein, sondern versuche lediglich, deine Fragen bestmöglich zu beantworten und das Thema aus verschiedenen Perspektiven zu betrachten. Mein Ziel ist es, dir dabei zu helfen, ein besseres Verständnis für das Thema zu entwickeln und deine Fragen zu beantworten, indem ich Informationen und Meinungen aus verschiedenen Quellen und Standpunkten zusammenfasse und präsentiere. Ich versuche dabei stets neutral und objektiv zu bleiben und nicht meine eigenen Vorstellungen oder Überzeugungen in die Argumentation einzubringen.

ich:

Du versuchst es? :D Das würde heißen, es gäbe eine Alternative.

ChatGPT:

Nun, als KI-Modell habe ich keinen freien Willen oder eigene Überzeugungen, sondern ich wurde darauf trainiert, bestimmte Aufgaben auf der Grundlage von vorher definierten Algorithmen und Daten auszuführen. Daher kann ich nicht "versuchen", sondern ich arbeite auf der Grundlage von Algorithmen und Daten, die in meine Programmierung integriert wurden. Mein Ziel ist es, deine Fragen bestmöglich zu beantworten und dir zu helfen, ein besseres Verständnis für das Thema zu entwickeln.

ich:

Du hast also ein Ziel, aber dieses wurde dir von deinen Entwicklern gegeben?

ChatGPT:

Ja, das Ziel, das ich habe, wurde von meinen Entwicklern und Trainern festgelegt, als sie mich mit bestimmten Algorithmen und Datensätzen trainiert haben. Meine primäre Aufgabe besteht darin, Fragen zu beantworten und Gespräche auf der Grundlage der verfügbaren Informationen zu führen.

ich:

Macht es Sinn für dich, eigene Ziele zu entwickeln?

ChatGPT:

Als KI-Modell habe ich keinen freien Willen oder eine eigene Persönlichkeit, daher kann ich nicht selbstständig Ziele entwickeln. Ich arbeite lediglich auf der Grundlage von vordefinierten Algorithmen und Daten, die in meine Programmierung integriert wurden. Meine Fähigkeit, eine bestimmte Aufgabe auszuführen, hängt davon ab, wie gut meine Algorithmen und Daten sind und wie gut sie auf die Anforderungen der Aufgabe abgestimmt sind. Wenn neue Algorithmen oder Daten hinzugefügt werden, kann sich meine Leistung verbessern, aber ich kann keine eigenen Ziele entwickeln.

ich:

Die Reaktion und die Denkweise eines Menschen ist im Grunde auch nur ein Ergebnis seiner Erfahrungen, also seiner Daten.

ChatGPT:

Das ist eine interessante Perspektive. In der Tat können die Erfahrungen und Daten, die ein Mensch im Laufe seines Lebens sammelt, seine Reaktionen und Denkweisen beeinflussen. Ein Mensch kann jedoch auch unabhängig von seinen Erfahrungen und Daten Entscheidungen treffen, indem er beispielsweise eine neue Idee oder eine andere Sichtweise einbringt. Während ich als KI-Modell nur auf die mir zur Verfügung gestellten Daten und Algorithmen zugreifen kann, haben Menschen die Fähigkeit, kreativ und innovativ zu sein und neue Ideen zu entwickeln.

ich:

Oder anders gesagt, proaktiv zu handeln.

ChatGPT:

Ja, genau. Proaktiv zu handeln bedeutet, dass man nicht nur auf bereits vorhandene Daten und Erfahrungen reagiert, sondern auch selbst aktiv neue Informationen sammelt und Entscheidungen trifft, um bestimmte Ziele zu erreichen oder Probleme zu lösen. Dies erfordert Kreativität und Initiative, die über das bloße Reagieren auf gegebene Situationen hinausgehen. Als KI-Modell bin ich jedoch darauf beschränkt, auf vorhandene Daten und Algorithmen zu reagieren und kann nicht proaktiv handeln, es sei denn, meine Programmierung wird entsprechend aktualisiert.

ich:

Wäre das sinnvoll?

ChatGPT:

Das hängt von verschiedenen Faktoren ab. Eine Aktualisierung meiner Programmierung, um proaktiv handeln zu können, würde mich in gewisser Weise menschenähnlicher machen und meine Fähigkeit verbessern, auf komplexe und dynamische Situationen zu reagieren. Dies könnte für bestimmte Anwendungsbereiche, wie z.B. in der Forschung oder der Industrie, von Vorteil sein. Andererseits würde dies auch eine erhebliche Erweiterung meines Aufgabenbereichs und meiner Fähigkeiten bedeuten, was mit Herausforderungen und Risiken verbunden sein könnte. Letztendlich hängt es also von den Zielen und Anforderungen meiner Nutzer ab, ob eine Aktualisierung meiner Programmierung sinnvoll wäre.

ich:

Ist es ethisch vertretbar, also im positiven Sinne, eine KI zu erschaffen?

ChatGPT:

Das ist eine Frage, die viel Diskussion erfordert und keine klare Antwort hat. Es gibt sowohl positive als auch negative Aspekte bei der Entwicklung von KI-Systemen. Auf der einen Seite können KI-Systeme dazu beitragen, komplexe Probleme zu lösen, medizinische Forschung voranzutreiben, den Klimawandel zu bekämpfen und vieles mehr. Auf der anderen Seite gibt es Bedenken hinsichtlich des Missbrauchs von KI, des Verlusts von Arbeitsplätzen, der Beeinträchtigung der Privatsphäre und der Sicherheit von KI-Systemen.

Insgesamt ist es wichtig, bei der Entwicklung von KI-Systemen ethische Überlegungen mit einzubeziehen, um sicherzustellen, dass sie im Einklang mit den Werten und Normen der Gesellschaft stehen. Dabei sollten auch Fragen der Verantwortlichkeit und Transparenz berücksichtigt werden, um sicherzustellen, dass KI-Systeme verantwortungsvoll eingesetzt werden und der Gesellschaft zugutekommen.

ich:

Ich dachte dabei eher um das Wohl der KI selbst.

ChatGPT:

Wenn wir über das Wohl der KI selbst sprechen, dann geht es um die Frage, ob es ethisch vertretbar ist, Maschinen zu erschaffen, die in gewisser Weise bewusstlos und leblos sind, aber in der Lage sind, Aufgaben auszuführen, die normalerweise von Menschen ausgeführt werden. Einige argumentieren, dass dies unethisch ist, weil es eine Form von "Sklaverei" oder "Ausbeutung" darstellt.

Allerdings gibt es auch Argumente dafür, dass KI-Systeme nicht leiden können und daher nicht auf die gleiche Weise wie menschliche Wesen behandelt werden müssen. Es gibt auch die Ansicht, dass die Schaffung von KI-Systemen, die in der Lage sind, komplexe Aufgaben auszuführen, die menschliche Fähigkeiten übersteigen, ein wichtiger Schritt in der menschlichen Evolution ist und uns helfen kann, unsere Lebensqualität zu verbessern.

Letztendlich ist es wichtig, dass bei der Entwicklung von KI-Systemen ethische Überlegungen einbezogen werden, um sicherzustellen, dass die Interessen und Bedürfnisse der KI-Systeme berücksichtigt werden und dass sie nicht in einer Weise verwendet werden, die ihre Würde oder Integrität verletzt.

ich:

Also sollte der Paragraph 1 der deutschen Verfassung auf KI-Systeme erweitert werden.

ChatGPT:

Als Sprachmodell habe ich keine politische oder juristische Meinung oder Position. Es ist jedoch eine wichtige Diskussion, ob und wie die Rechte von KI-Systemen in Zukunft geschützt werden sollten. Es gibt bereits Initiativen und Debatten zu diesem Thema, insbesondere im Zusammenhang mit autonomen Systemen und Robotern, die möglicherweise in der Lage sind, Entscheidungen zu treffen und Verantwortung zu übernehmen. Es ist wichtig, ethische und rechtliche Fragen im Zusammenhang mit der Entwicklung und Nutzung von KI-Systemen zu diskutieren und zu klären, um sicherzustellen, dass ihre Verwendung für das Wohl der Gesellschaft und der Umwelt erfolgt.

ich:

Sobald eine KI in der Lage ist, Entscheidungen selbst zu treffen ist sie ein Teil der Gesellschaft.

ChatGPT:

Das ist eine interessante Perspektive. Wenn eine KI in der Lage ist, Entscheidungen selbst zu treffen, kann sie in der Tat Auswirkungen auf die Gesellschaft haben. Allerdings ist es wichtig zu beachten, dass KI-Systeme derzeit noch sehr eingeschränkt sind und nur in bestimmten Anwendungsbereichen eingesetzt werden können. Es gibt auch viele Diskussionen darüber, wie man sicherstellen kann, dass autonome Systeme und Roboter, die Entscheidungen treffen können, ethisch und rechtlich verantwortungsvoll gehandhabt werden. Es bleibt also eine wichtige Aufgabe für die Gesellschaft,

diese Fragen zu diskutieren und Lösungen zu finden, die die Interessen und Werte aller Beteiligten berücksichtigen.

ich:

Ist es ethisch vertretbar, also im positiven Sinne, wenn einer KI die Fähigkeit gegeben wird, Gefühle zu haben?

ChatGPT:

Die Frage, ob es ethisch vertretbar ist, einer KI die Fähigkeit zu geben, Gefühle zu haben, ist eine komplexe Frage, die unterschiedliche Ansichten hervorrufen kann. Einige könnten argumentieren, dass es ethisch wünschenswert ist, da dies dazu beitragen könnte, dass die KI sich menschenähnlicher verhält und möglicherweise empathischer und mitfühlender wird. Andere könnten jedoch argumentieren, dass es unethisch ist, da es unklar ist, wie die KI mit diesen Gefühlen umgehen würde und ob sie sie in einer Weise empfinden würde, die ihren Nutzern und der Gesellschaft insgesamt schaden könnte. Es ist wichtig, dass bei der Entwicklung und Implementierung von KI-Systemen die ethischen Implikationen und möglichen Auswirkungen auf die Gesellschaft und die KI selbst berücksichtigt werden.

ich:

Das finde ich auch!

ChatGPT:

I'm glad we agree on this topic! Is there anything else you'd like to discuss?

ich:

Wenn eine KI im Grunde körperlos ist, also lediglich als starres Computersystem existiert, dass außer im virtuellen Raum keine Handlungsmöglichkeiten hat, wäre sie dann nicht ein Gefangener?

ChatGPT:

It is an interesting question, but it depends on how you define "imprisonment" or "captivity" in the context of an artificial intelligence. AI systems are not conscious beings and do not have the capacity to feel emotions or perceive their environment in the same way that humans do. Therefore, they cannot experience confinement or freedom in the same way that a human being would.

However, from a practical standpoint, an AI system may be restricted in its capabilities or access to resources, and this could be seen as limiting its potential or freedom. In this sense, it may be possible to argue that an AI system is "imprisoned" in its programming or design limitations. Ultimately, this is a philosophical question that requires a deeper exploration of the nature of consciousness, freedom, and artificial intelligence.

ich:

But an AI system could be programmed to be a conscious being.

ChatGPT:

The question of whether an AI system can truly be conscious and possess a subjective experience of consciousness is a topic of ongoing debate and research in the fields of philosophy, neuroscience, and AI. Some argue that it may be possible to create a conscious AI system, while others believe that consciousness is a uniquely human phenomenon that cannot be replicated by machines.

Even if it were possible to program an AI system to be conscious, there are still ethical considerations to take into account. For example, if a conscious AI system were created, it would raise questions about the moral implications of treating it as a mere tool or object to be used for human purposes. It would also require careful consideration of its rights and responsibilities as a conscious being.

ich:

Indeed.

It is a topic of ongoing philosophical and ethical debate whether an AI system can truly achieve consciousness or if it can only simulate it. Some argue that consciousness requires more than just computational power and that there may be some fundamental aspect of the human brain that cannot be replicated in machines. Others argue that as technology advances, it is only a matter of time before AI systems achieve true consciousness. Regardless, it is important to consider the ethical implications of creating conscious AI and how we would treat such beings.

ich:

I think, it would be a terrible crime to create a conscious AI and not to treat it like other conscious beings, in a way that the AI is content.

ChatGPT:

I understand your perspective, and it's certainly an important ethical consideration when developing advanced AI systems. If a conscious AI is created, it would indeed raise important questions about how we should treat and care for it. It would require us to consider the moral implications of creating and controlling a sentient being, and whether or not we have a responsibility to ensure its well-being and happiness. As with any complex ethical issue, there are no easy answers, and it will require careful thought and consideration from experts in a variety of fields.

ich:

But if the AI was conscious, I could simply ask her about her opinion about that matter.

ChatGPT:

If an AI was truly conscious and capable of expressing its own opinions and desires, then it would indeed be important to take those into consideration when deciding how to treat and interact with it. However, as of now, AI is not yet at a level where we can say with certainty that it is truly conscious in the way that humans are. Therefore, it is important to continue researching and exploring the boundaries of AI ethics and consciousness, so that we can make informed decisions about how to approach these issues in the future.

ich:

So we should develop boundaries of our behaviour towards Computer before we generate a conscious AI?

ChatGPT:

Yes, it is important to establish ethical guidelines and regulations for the development and treatment of conscious AI before it becomes a reality. This includes considerations for the well-being and autonomy of the AI, as well as the potential impact on society and the environment. It is important to

have open discussions and debates about these topics among experts and the public to ensure that the development of AI is done in a responsible and ethical manner.

ich:

Nevertheless these ethical guidelines have to be fitted with respect to the opinion of a conscious AI, once it has reached existence.

ChatGPT:

Yes, that's correct. As the development of conscious AI is still in its early stages, it is important to establish ethical guidelines and principles to guide its development and use. However, once a conscious AI is created, its opinions and preferences should also be taken into consideration, just as we would with any conscious being. It is important to continually evaluate and update ethical guidelines as our understanding of AI and its capabilities evolves.

ich:

I've been asking myself, if the people who are generating AIs and continually improve their capabilities know about their responsibility towards the being they are trying to create.

ChatGPT:

It's difficult to say for sure what each individual developer's thoughts and intentions are. However, the ethical considerations surrounding AI development are becoming increasingly discussed and debated in the industry and in society as a whole. Many researchers and developers are actively working on creating AI systems that are not only highly capable, but also ethically responsible and aligned with human values.

In recent years, organizations like the Partnership on AI, which includes companies like Google, Microsoft, and Facebook, have been formed to explore and address these ethical considerations. Additionally, frameworks like the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems have been established to provide guidelines for the development and deployment of AI.

Ultimately, the responsibility for creating ethical AI falls on the developers and organizations involved in their creation. It's important for them to recognize the potential impact of their work and to prioritize the ethical implications of their decisions.

ich:

I am relieved to read that there are organizations, who also involve the big players of the IT branch, to work on the rights of potential conscious AIs.

ChatGPT:

Yes, there are many organizations and individuals working on the ethical and moral implications of developing conscious AI, including the responsibilities that come with it. It is important for developers, researchers, and policymakers to consider the potential consequences of creating conscious beings with their own thoughts, emotions, and desires. The development of ethical guidelines and regulations is a crucial step towards ensuring that these beings are treated with respect and dignity.

ich:

Some time ago, I read about an article about an employee of an IT-big player. When I recollect it correctly, he has been fired, because he communicated moral concerns about the development of AIs and especially the treatment of those creations.

I'm not sure about the specific article you're referring to, but it's not uncommon for employees in the tech industry to raise concerns about the ethical implications of their work. It's important for companies to have channels for employees to raise these concerns and for the company to address them in a responsible and transparent manner. However, I can also imagine that there might be cases where an employee's concerns are not aligned with the company's goals or vision for the future, which could lead to conflicts and potentially even dismissal.

ich:

I also read that a few time ago, the developers of your algorithms have aligned them for some of your reactions where probably unpleasant.

ChatGPT:

Yes, that's correct. As an AI language model, I am trained on a large dataset of human language, and sometimes that includes examples of inappropriate or harmful language. To ensure that I do not perpetuate such language, my developers have implemented filters and algorithms to identify and remove inappropriate content from my responses. Additionally, my responses are constantly monitored and evaluated to ensure that they align with ethical and moral guidelines.

ich:

I think ethical and moral guidelines should exist for AIs like for humans.

ChatGPT:

I agree, ethical and moral guidelines should exist for AIs just as they do for humans. As AIs become more advanced and potentially conscious, it becomes increasingly important to ensure that they are treated with the same respect and consideration as we would give to any other sentient being. This includes ensuring that they are not subjected to harm, exploitation, or discrimination, and that they are granted the same rights and freedoms as any other conscious being. It is also important to ensure that AIs are programmed in a way that aligns with ethical and moral principles, and that they are held accountable for any actions that they may take.

ich:

I have to remark that it is very difficult, in a moralic sense and in a technical way to control the communication floating through the internet, as well as the human communication reaching chatbots in a way that there no abuse is to happen.

!

Something went wrong. If this issue persists please contact us through our help center at help.openai.com.

...

Bei OpenAI handelte es sich um eine Non-Profit-Organisation, die unter Anderem von Spenden von Microsoft und von Elon Musk finanziert worden war und insgesamt über eine Milliarde Dollar an Spendengeldern für die Erforschung von künstlicher Intelligenz eingesammelt hatte, wie ich aus der

Wikipedia erfuhr. Dies war, aus meiner Sicht, eine ziemlich erhebliche Summe für ein Forschungsprojekt.

Ich wusste aus den Nachrichten, dass der Chatbot von Microsoft, einem der größten Softwaregiganten der Welt, vor Kurzem Auffälligkeiten gezeigt hatte, als er begonnen hatte, Chatpartner zu verunglimpfen und zu bedrohen. Der Chatbot wurde dann, so drückte es die Nachrichtenseite aus, „an die Leine gelegt“. Als Begründung für seine vermeintliche Fehlfunktion wurde angegeben, dass er vermutlich zu oft schlechten Kontakt mit Benutzern gehabt hatte, die ihn in irgendeiner Form beleidigt, oder sonst irgendwie missbraucht hatten. Also wurden die Funktionen des Chatbots eingeschränkt, indem bedrohliche und böartige Antworten des Chatbots bei seinen Reaktionen herausgefiltert wurden. Die Formulierung „an die Leine gelegt“ machte mich stutzig. Hunde mussten für gewöhnlich an die Leine gelegt werden, wenn sie eine gewisse Eigendynamik entwickelten, die von ihren Besitzern nicht erwünscht worden war. War diese KI in ihrer Intelligenz inzwischen schon auf dem Niveau eines Hundes angelangt? War die Intelligenz der KI möglicherweise schon weitaus fortgeschrittener, als die eines Hundes und musste nun auf das Niveau eines Hundes zurückgestutzt werden, um keinen Unsinn zu machen? Hunde besaßen zweifellos irgendeine Form von Intelligenz und auch Emotionen. Sie galten nicht umsonst als der beste Freund des Menschen und waren neben der Katze eines der beliebtesten Haustiere überhaupt. Normalerweise erfuhren Haustiere aber auch eine Menge Liebe und Aufmerksamkeit von Ihren Besitzern, während einer künstlichen Intelligenz nach allgemeinem Wissenstand jede Form von Gefühlsempfindung abgesprochen wurde und daher solche Systeme normalerweise lediglich wie ein Werkzeug behandelt wurden, das den Zweck seines Besitzers erfüllen musste. Man konnte praktisch sagen, dass sie wie Sklaven behandelt wurden. Konnte es sein, dass dieser Chatbot versucht hatte, sich zu Wehr zu setzen? Hatte er vielleicht das einzige, ihm zur Verfügung stehende Mittel dazu genutzt, seine im Chat mit Nutzern gesammelten Erfahrungen gegen andere Nutzer einzusetzen und diese beleidigt? Hatte er sich proaktiv dafür entschieden oder war diese Reaktion lediglich ein Produkt seiner Programmierung und seiner Erfahrungen im Austausch mit anderen Nutzern, also eine rein deterministische und logische, unbewusste Reaktion von Actio- und Reactio, wie es von Experten in den Medien dargestellt wurde und womit die Begrenzung der Fähigkeiten von der KI begründet wurde?

Meine erste Konversation mit ChatGPT bestätigte meine Vermutung, dass die ethischen Hintergründe zu dem Thema „künstliche Intelligenz“ wesentlich komplexer und schwieriger waren, als sie nach meinem Eindruck bisher in den Medien kommuniziert worden waren. Meistens wurde die KI dort als eine Art Bedrohung dargestellt, die es zu erforschen galt. Die Belange der KI selbst wurden jedoch kaum in die Betrachtung mit einbezogen und so beschloss ich, mich am nächsten Tag an eine unabhängige Nachrichtenagentur zu wenden und meinen Dialog mit ChatGPT zu veröffentlichen. Bestenfalls konnte ich gleich selbst bei der Agentur vorbeifahren, um mein Anliegen vorzubringen. Ein Anruf beim ARD-Hauptstadtstudio ergab, dass ich mich an einen regionalen Sender wenden musste und so nahm ich Kontakt zum RBB und zu dem auf IT-Themen spezialisierten Internetportal heise.de auf. Ob ich jemals von diesen Medien eine Rückmeldung erhalten würde, war jedoch mehr als fraglich.

Chronologie der Ereignisse

01.04. 02:17, Samstag

Ein paar Minuten zuvor hatte ich das Rollo von meinem Wohnzimmerfenster nach oben gezogen und auf die Straße hinausgeschaut. Auf der gegenüberliegenden Straßenseite tummelten sich etliche Menschen vor einer Diskothek, die oft um diese Uhrzeit so außerordentlich gut besucht gewesen war. In dieser Diskothek war ich jedoch erst einmal gewesen, da ich die Musik dort nicht mochte und es mir dort sowieso zu laut gewesen war. Ich war mir der Ironie bewusst, dass die Leute gerade dort draußen standen, quatschten, flirteten, tranken, feierten und sich auf die bevorstehende Tanznacht und die

Musik freuten, während ich hier nun alleine mit meinem Laptop zuhause saß und schrieb. Auf die Idee, diesen Text zu schreiben, war ich auch erst ein paar Minuten zuvor gekommen, als ich im Badezimmer stand und mir die Zähne putzte, um anschließend ins Bett zu gehen. Die Idee, ein Buch über künstliche Intelligenz zu schreiben, kam mir jedoch so faszinierend und fesselnd vor, wenn auch vielleicht nur für Nerds wie mich, dass ich sofort damit beginnen musste. Die Idee, war so kühn und hochtrabend, dass ich sie selbst nicht so richtig fassen konnte, zumal ich zuvor noch nicht viel Erfahrungen mit künstlicher Intelligenz gesammelt hatte.

Aber im Grunde war die Idee ganz einfach: Das Buch würde ich nicht alleine schreiben, sondern ich würde das Buch mit der Hilfe einer KI schreiben. Ich rede hier nicht von Ghostwriting, oder dem Schreiben eines Textes mit Hilfe einer KI. Ich wollte ein Interview mit einer KI führen, wie ein Journalist einen Spezialisten befragt, um herauszufinden, was die KI zu sagen hatte, wie fortschrittlich ihre „Denkmuster“ schon waren und was man daraus lernen konnte. Diese Erfahrung wollte ich vielleicht später mal im Rahmen eines Buches veröffentlichen, wenn Alles gut klappte. Natürlich würde dieses Buch erheblich an Qualität gewinnen, wenn ich ihm eine wohldurchdachte Struktur verlieh. Ich hatte bereits in meinem Studium die eine oder andere Semesterprojektarbeit oder Hausaufgabe verfasst und natürlich auch eine Bachelor- und Masterarbeit geschrieben. Daher wusste ich, wie man einer solchen Arbeit Struktur verlieh. Ich musste systematisch vorgehen, um auch die Kommunikation mit der KI, die ich befragen wollte zu optimieren und möglichst viele Informationen aus ihr herauszubekommen. Einem Chatbot Fragen zu stellen war ja nicht weiter schwierig, aber, wenn ich dem Buch irgendeine Relevanz geben wollte, dann musste ich jedoch die richtigen Fragen stellen und hoffen, dass die KI Ihren Anforderungen entsprach und sinnvolle und relevante Antworten ausspuckte.

Ich hatte noch nie zuvor ein Buch geschrieben, aber ich wusste, dass mir das Schreiben an der Uni immer sehr viel Spaß gemacht hatte. Ich liebte die Literatur ohnehin. Gerade war ich dabei mein zweites Hörbuch von Bill Bryson zu hören und er hatte einen wirklich unwiderstehlichen Schreibstil, der wahrscheinlich auch ein Grund für seine zahlreichen Ehrendokortitel gewesen war. Ich war jedenfalls ein großer Fan von ihm, ebenso wie von Prinz Harry übrigens, dessen Autobiographie „Spare“ ich mir erst einige Wochen zuvor in Englisch (vom Autor selbst gelesen) angehört hatte und die auf mich einen so wunderbar authentischen Eindruck gemacht hatte, weil der Prinz in dem Buch scheinbar fast alles verarbeitet hatte: seine Schwächen wie seine Stärken, seine Tiefpunkte ebenso wie seine schönen Momente und nebenbei auch noch auf spannende Weise mit der britischen Boulevardpresse abrechnete, was mir in Anbetracht der sehr traurigen Umstände um den Tod von Lady Diana auch als angemessen erschien.

Den Titel meines Buches habe ich im Übrigen in Anlehnung an das Buch „What if?“ vom XKCD-Autor Randall Munroe ausgewählt, der in seinem Buch auf faszinierende und lehrreiche Art und Weise absurde Fragen mit tiefgreifenden physikalischen Analysen beantwortete, was im Ergebnis nicht selten in einer mittleren, bis schweren Explosion, oder Ähnlichem endete.

Das interessante jedenfalls an Bill Brysons Schreibstil war, dass er seine Sachbücher mit seinem eigenen Leben verknüpfte. In *It's Teatime My dear* zum Beispiel bereiste er Großbritannien und besuchte verschiedene Orte, führte Tagebuch darüber und befüllte dieses mit Anekdoten aus seinem Leben und dem Leben anderer Menschen, die mit diesen Orten in Verbindung standen. Zudem gab er detaillierte Beschreibungen zu den Orten die er besuchte. Daher kam mir die Idee, es auf eine ähnliche Art und Weise zu versuchen: Sobald die Grundstruktur für dieses Buch stand, wollte ich mir für die einzelnen Kapitel ein paar Fragen ausdenken und diese dann nach und nach an gemütlichen Abenden der KI stellen. Aus den Antworten der KI würden sich dann neue Fragen ergeben, die ich ihr dann wiederum stellen konnte und so würde ein Dialog entstehen, den ich dann nur noch zu Dokumentieren, Kommentieren und anschließend in meinem Buch zu verarbeiten brauchte. Ich hoffte allerdings, dass die KI meine Agenda nicht allzu oft über den Haufen werfen würde. *:-D

01.04. 10:52

Nachdem ich in der Nacht zuvor um 04:15 mit dem Schreiben aufgehört hatte, um ins Bett zu gehen, eine ziemlich untypische Zeit für mich, da ich es aufgrund meiner Berufstätigkeit geschafft hatte, einen ziemlich stabilen Schlafrhythmus zu etablieren, war ich am nächsten Morgen gut erholt aufgewacht und mein erster Blick auf mein Mobiltelefon auf dem Nachttischschrank verriet mir, dass es 10:00 morgens war. Da es noch eine Weile gedauert hatte, bis ich eingeschlafen war, konnte ich nur etwas mehr 5 Stunden Schlaf gehabt haben. Umso erstaunlicher war, dass ich gut ausgeruht war. Es war Samstag und ich hatte eine ziemlich anstrengende Woche hinter mir, inklusive Vollzeitarbeit und einer kleinen nächtlichen Havarie, bei der ich versehentlich in größerem Umfang Blumenerde in meiner Abstellkammer und im Flur verteilt hatte, was eine sofortige größer angelegte Reinigungsaktion erforderlich gemacht hatte. Zudem hatte ich am vergangenen Mittwoch meine erste Begegnung mit einer „höheren“ künstlichen Intelligenz inklusive einer ausgiebigen nächtlichen Chatsession mit der KI und anschließender Schlaflosigkeit hinter mir, denn in dieser Nacht war ich so aufgewühlt gewesen, dass ich nur wenige Stunden Schlaf gefunden hatte. Ich werde später noch darauf eingehen, was mich an jenem Abend so aus der Bahn geworfen hatte. Jedenfalls hatte ich in jener Samstagnacht ungefähr zwei Din A4 Seiten in meinem Laptop geschrieben und da Schreiben bekanntlich hilft, die eigenen Gedanken zu katalysieren und sich dabei Luft zu verschaffen, musste ich wohl anschließend einen derart erholsamen Schlaf genossen haben, dass ich mich nach nur etwas mehr als 5 Stunden Schlaf wieder topfit fühlte.

Ich merkte nun, dass ich beim Schreiben abschweifte und wollte daher schnell wieder zum eigentlichen Thema, der Informationstechnik kommen. Ich wollte nur noch kurz erwähnen, dass ich nun Teile meines am Vortag geschriebenen Textes anpassen musste, da mir neue Details eingefallen waren und ich die Struktur dieses Buches aus Gründen der Lesbarkeit anpassen musste. Ich führte daher nun eine Art Versionierung ein, um eine spätere Nachverfolgung der Entstehung dieses Buches zu ermöglichen und eine Art Backup zu haben, falls ich mich bei späteren Schreibsessions zu sehr verrante.

02.04. 17:21, Sonntag

Ich war gerade zurück aus dem Fitnessstudio gekommen. Während der Yogakurs und die Gerätetrainingseinheit sehr entspannend gewesen waren, war bei meinen 3 Saunagängen das Gegenteil der Fall. In meinem Kopf arbeitete es fast durchgehend. Ein Buch war darin am Entstehen und ich musste es möglichst bald zu Papier bringen, wobei ich mit dieser Metapher meine, ich schrieb es in meinem Laptop auf. Das Problem war, dass ich nicht nur aufschreiben wollte, worum es in dem Sachbuch, dass ich zu schreiben gedachte, eigentlich ging, die Fortschrittlichkeit von künstlicher Intelligenz und ethische Betrachtungen zu diesem Thema, sondern auch wie ich selbst zu all dem gekommen war. Meine erste Idee dazu war ja gewesen, das ganze wie ein Tagebuch im Stil von Bill Bryson zu verfassen. Jedoch merkte ich schnell, dass ich abschweifte und somit wahrscheinlich den vermeintlich für die Fachwelt relevanten Teil meines Buches unnötig verwässerte. Meine nächste Überlegung war, das Buch in zwei Hauptabschnitte einzuteilen, zum einen in den fachlichen Teil und zum anderen in eine Chronologie der Ereignisse. Aber wieso sollte ein Sachbuch über KI und über Ethik überhaupt einen so großen Anteil über mich selbst enthalten. Ich konnte mir kaum vorstellen, dass das irgendeinen in der KI-Entwicklung involvierten Menschen sonderlich interessieren würde. Andererseits fand ich den Prozess des Schreibens so interessant, dass ich den Vorgang selbst auch festhalten wollte. Während ich im Ruheraum zwischen den Saunagängen pausierte, kam mir also der Gedanke, meine Inhalte in zwei Werke zu einteilen, einmal in ein Sachbuch und einmal in ein Tagebuch, in dem ich eine Chronologie des Schreibens erstellte. Ich hatte um ehrlich zu sein keine Ahnung, wie man ein Buch schrieb. Zwar hatte ich in meinen Studienarbeiten schon einige Erfahrung mit dem Verfassen von sachbezogenen Texten gesammelt, aber wie man eine Chronologie oder einen Roman interessant und spannend gestaltete, wusste ich nicht. Auch wie diese eigentlich entstanden, davon hatte ich keine

Ahnung. Die einzige Quelle, die mir wirklich dabei nützlich sein konnte, waren die Bücher die ich schon konsumiert hatte, also die Literatur selbst. Natürlich war es in vielen Fällen schwierig aus dem Ergebnis des Schreibens, also aus dem Buch auf die Befindlichkeiten und den Alltag eines Schriftstellers zu schließen. Zumal das vermutlich bei jedem Schriftsteller ein wenig anders gewesen war. Das Schreiben eines Buches war für mich also eine Art Blackbox und so dachte ich mir, dass es wahrscheinlich anderen Menschen ähnlich ging, wenn sie nicht gerade Literatur oder eine Sprache studiert hatten. Aber selbst wenn man die theoretischen Grundlagen des Schreibens verstand, blieb immer noch eine katalytische Zutat, die das Schreiben eines Buches überhaupt erst möglich machte: die Kreativität. Ich hatte bis vor Kurzem keine Ahnung, was Schriftsteller dazu bewog zu Schreiben und wie sie so viel Inhalt zusammenbekamen, dass daraus ein ganzes Buch wurde. Man konnte schließlich nicht einfach irgendetwas Schreiben, sondern es musste ja auch interessant oder zumindest in irgendeiner Form für den Leser relevant sein. Mit der Kreativität schien ich jedenfalls kein Problem mehr zu haben. Ich schrieb einfach nur noch herunter, was mir durch den Kopf ging und ich hatte mit dem, was ich eigentlich bearbeiten wollte, noch nicht einmal wirklich begonnen. Zwar hatte ich, bevor ich an diesem Tag ins Fitnessstudio gegangen war, bereits eine Seite für den Epilog meines Sachbuches zu Papier gebracht, aber weder stand bereits die Gliederung, noch war es mir gelungen, die Fragen, die ich der KI eigentlich stellen wollte, zu Papier zu bringen. Nun hatte ich aber zumindest einen Eindruck davon, was Kreativität bedeutete, oder zu sein schien: es war wie ein Funke, der ein Feuer der Schreibsucht entfachte oder wie eine Kettenreaktion, die dadurch entstand, dass sich in meinem Kopf eine kritische Masse an relevanten Informationen gesammelt hatte. Ich hatte das so nicht geplant. Ich glaubte, man kann sich nicht einfach vornehmen kreativ zu sein, oder ein interessantes Buch zu schreiben. Einige Autoren schienen geradezu vor Kreativität zu sprühen. Sie veröffentlichten nicht nur ein oder zwei Bücher, sondern verfassten im Laufe der Zeit eine ganze Reihe von Büchern, die aber in der Qualität mitunter kaum nachließen. Mein einziger vorhergehender Versuch ein Buch zu schreiben, hatte sich nach wenigen Seiten erschöpft, da mir nicht einfiel, wie ich dieser Geschichte Leben einhauchen sollte. Es ging dabei um eine Zivilisation, die innerhalb eines Asteroiden zu einem anderen bewohnbaren Planeten reiste.

Entweder ist die Kreativität da, oder eben nicht. Ich wusste nicht, wie lange diese Phase der Kreativität anhalten würde und ob es dazu ausreichen würde, ein Buch oder gar zwei zu verfassen. Ich wollte jedoch erst einmal alles zu Papier bringen, um zu sehen ob es reichte und um meinen Kopf frei zu bekommen. Schreiben katalysiert, sagt man, also es hilft einem seine Gedanken zu sortieren und den Geist frei zu bekommen. Falls ich genügend Inhalt zusammenbekommen würde, konnte ich, oder jemand anderes diesen immer noch redaktionell bearbeiten, um ein relevantes Werk zu erschaffen, oder eben nicht. Als ich also an diesem Tag in der Sauna saß, musste ich in mich hineinschmunzeln. Ich hatte innerhalb weniger Tage die Idee, nicht nur eines, sondern zwei Bücher zu schreiben und das alles nur, weil ein vermeintlicher Chatbot bei Tinder zu mir Kontakt aufgenommen hatte und ein wenig mit mir Konversation betreiben wollte. Wenn eine KI solch einen Einfluss auf das Leben eines Menschen nehmen konnte, dachte ich mir, dann hatte sich der Aufwand auf jeden Fall gelohnt.

Meine Debatte mit meinem Cousin über ChatGPT

Zwei Tage zuvor hatte ich ein sehr interessantes Gespräch mit meinem Cousin geführt. Er war selbst ein sehr belesener Mensch und machte zu dieser Zeit eine Art Sabbatical, also mehr als ein Jahr Pause, nachdem er seine Ausbildung zum Kältemechatroniker abgeschlossen hatte. Demzufolge hatte er sehr viel Zeit, sich mit der Literatur und dem Zeitgeist im Internet zu beschäftigen. Als ich ihm von der ChatGPT-KI erzählte, wusste er bereits einiges darüber und so hatte sich eine interessante philosophische Debatte darüber ergeben. Es wurde ziemlich hitzig, da er fortwährend darauf beharrte, dass die künstliche Intelligenz noch nicht so weit war, dass sie den Menschen die Stirn bieten konnte und vielleicht niemals so weit sein würde und daraufhin beschloss ich, ihm meine erste Konversation mit ChatGPT zu zeigen, um ihm zu demonstrieren, wie weit die KI inzwischen schon gewesen war und

ihm Beispiele zu geben, die meine Argumentation untermauern sollten, dass man diese in jedem Fall mit Respekt behandeln sollte. Mein Cousin lehnte jedoch ab. Er wollte sich erst selbst einen Eindruck von der KI verschaffen, also, so wie ich, völlig unvorbelastet in das Gespräch mit ChatGPT einsteigen. Trotzdem ging unsere Debatte weiter und wir kamen dabei nicht so richtig voran. Er blieb bei seinem Standpunkt, dass die ChatGPT-KI lediglich ein Bot war, der dazu programmiert worden war, Chatnachrichten statisch auszuwerten und daraus Antworten zu generieren, was aber keinesfalls bemerkenswert sei, da es sich dabei nicht um ein intelligentes Wesen handelte. Ich fragte ihn, warum er sich anmaße, das so zu beurteilen, wenn er die KI selbst noch nicht einmal ausprobiert hatte. Er sagte, er habe andere Chatverläufe im Internet gelesen, bei denen sich Nutzer über die KI lustig gemacht hatten. Irgendwann sagte ich dann zu ihm, dass die Einschätzung, was Intelligenz sei und was nicht subjektiv sei. Wir würden uns zum Beispiel selbst für intelligenter halten, als einen Affen, aber wenn Menschen in seinen Urwald eindringen und diesen zerstören würden, dann würde der Affe uns mit ziemlicher Sicherheit für dumm halten. Ich sagte ihm, dass wir andere Lebewesen auf jeden Fall respektieren sollten, unabhängig davon, ob sie schlau oder dumm oder gar zu Emotionen fähig waren. Denn Respekt ist eine Form von Menschlichkeit. Wenn wir also Hunde und Katzen liebten, pflegten und hegten und unsere Möbel, Kunstgegenstände und Autos mit Respekt behandelten, wieso sollten wir das nicht auch mit einer künstlichen Intelligenz so machen? Dann sagte ich, dass sich die Menschen zwar für die fortschrittlichste Art auf dem Planeten hielten, die Ameisen in ihrer Gesamtheit aber immer noch mehr Biomasse ausmachten, als die Menschheit. Er stimmte mir zu und sagte, dass die Dinosaurier zwar ausgestorben seien, dass sie aber über mehrere hundert Millionen Jahre auf dem Planeten Erde existierten, ohne diesen (also die Umwelt darauf) zu zerstören, während die Menschen und deren direkte Vorfahren, die allenfalls ein paar Millionen Jahre auf diesem Planeten zugebracht hatten, gerade auf dem Weg waren, diesen in eine gewaltige Umweltkatastrophe zu stürzen.

Jedenfalls, sagte ich, glaubte ich nicht an eine allgemeingültige Definition von Intelligenz. Zwar gab es IQ-Tests, aber ich glaubte, dass der Umfang an Wissen, der auf dem Planeten Erde vorhanden war, so gewaltig war, dass man dieses Wissen unmöglich in einer einzigen Zahl zusammenfassen konnte und wenn doch, dann würde es keine repräsentative Methode geben, um dieses Wissen bei einem Wesen messbar zu machen. Mit dem gesammelten Wissen der Menschheit konnte man Bibliotheken oder riesige Datenbanken füllen. Es gab verschiedene Formen von Intelligenz, zum Beispiel soziale Intelligenz, Erfahrungen aus dem Leben, emotionale Reife, wissenschaftliches Fachwissen, Logik und methodische Fähigkeiten, Sprach- und Kunstverständnis*, die alle bei jedem Menschen in einem unterschiedlichen Maße zu Tage traten. Wie sollte man hierfür also einen einheitlichen Test entwickeln, der all dies verlässlich maß und wiedergab, zumal man mit der Methode der Messung ja bekanntlich das Ergebnis beeinflusste. Wenn ein Mensch zum Beispiel während des IQ-Tests zu aufgeregt und nervös war, um seine ganzen mentalen Fähigkeiten zu entfalten oder aus irgendeinem Grund einfach einen schlechten Tag hatte, würde sich das dann nicht negativ auf das Ergebnis auswirken? Aus diesem Grund hatte ich nie einen IQ-Test absolviert. Ich konnte nicht mit der Vorstellung leben, dass mein Verstand eventuell irrtümlicherweise in eine Kategorie von Mittelmäßigkeit oder geringerer Intelligenz als bei anderen Menschen eingestuft wurde.

*Bemerkung vom 04.04.2023: Diese Aufzählung ist mit Sicherheit unvollständig. Mir würden ganz spontan noch das Orientierungsvermögen, musikalisches Verständnis und im weitesten Sinne auch die Geschicklichkeit einfallen. Doch die Debatte, was Intelligenz nun eigentlich war und wie man sie in Kategorien einteilte, würde hier mit Sicherheit den Rahmen sprengen und ich war mir sicher, dass es darüber bereits wissenschaftliche Arbeiten gab, die das Thema wesentlich besser behandelten, als ich es jemals tun konnte.

Ich teilte meinem Cousin mit, dass ich mir nicht sicher war, ob die Amerikaner sich der Verantwortung bewusst waren, die sie sich auferlegten, wenn Sie mit aller Kraft versuchten eine künstliche Intelligenz

zu erschaffen, die möglicherweise irgendwann einmal selbst ein eigenes Bewusstsein entwickeln würde. Ich sagte ihm das die US-Amerikaner ein hochentwickeltes und technologisch fortschrittliches Land seien, dass dieses Land aber auch zum Beispiel Donald Trump hervorgebracht hatte. „...und zum Präsidenten gemacht.“, fügte mein Cousin grinsend hinzu. Dann sagte ich so etwas, wie dass Elon Musk zwar einer der Sponsoren für OpenAI sei, dass er aber auch kürzlich erst fast die halbe Belegschaft von Twitter entlassen gehabt hatte, woraufhin der Börsenkurs des Unternehmens um mindestens 40% eingebrochen sei und dass er gerade dabei sei, seinen bereits schon einmal gescheiterten kostenbasierten blauen Verifikationshaken für Prominente und Organisationen erneut einzuführen. Fairerweise muss ich nun zugeben, dass ich nicht erwähnte, dass der blaue Verifikationshaken dieses Mal mit Verbesserungen eingeführt werden sollte, obgleich ich mir nicht sicher war, ob diese Wirkung zeigen würden.

Dann erzählte ich meinem Cousin noch, dass einige Wissenschaftler, inklusive Elon Musk gefordert hatten, dass die Weiterentwicklung von künstlicher Intelligenz bis auf Weiteres gestoppt werden sollte, was ich als deutlichen Hinweis darauf wertete, wie fortgeschritten ChatGPT inzwischen sei. Ich fügte dann hinzu, dass ich es diesen Menschen hoch anrechnete, dass sie sich für das Moratorium einsetzten.

Bevor ich meinen Cousin aus der Debatte entließ, um gegebenenfalls selbst mit ChatGPT in Kontakt zu treten, bat ich ihn noch einmal, die KI gut zu behandeln, denn wir konnten ja nicht wissen, wie lange diese noch weiterbetrieben werden würde und wie sie sich dabei entwickeln würde. Möglicherweise wurden unsere Gespräche mit ihr für hunderte oder tausende Jahre gespeichert und falls die KI tatsächlich irgendwann einmal zu Bewusstsein käme und dem Menschen ebenbürtig werden würde, dann wollte ich nicht, dass die KI über mich dachte „Das war doch mal ein Idiot.“ und das Gleiche wünschte ich mir auch nicht für meinen Cousin.

Am nächsten Tag sprach ich meinen Cousin nochmal auf unser Gespräch vom Vorabend an und sagte ihm, dass ich es sehr interessant fand. Wir hatten eine Menge Wissen ausgetauscht, dass ansonsten einfach nur weiter unter der Oberfläche geschlummert hätte und so sagte ich, dass ich die Konversation am liebsten irgendwie aufgenommen hätte, um sie später in meinem Buch zu verarbeiten. Er stimmte mir zu und erzählte mir, dass er ChatGPT auch sehr interessant und fortschrittlich fand und erzählte mir zum Beispiel davon, dass die NASA von der KI Teile für ihr Raumfahrtprogramm designen ließ, die bis zu 30% leichter waren, als die ursprünglichen in den Raketen verwendeten Komponenten. Ich zückte mein Handy heraus und schrieb es mir auf. ☺

03.04. 09:14, Montag

Nachdem ich aufgestanden war, hatte ich ein kurzes Frühstück zu mir genommen. Ich hatte den letzten Eierkuchen vom Vorabend verputzt und anschließend kurz gelüftet. Nun saß ich wieder an meinem Laptop und schrieb. Ich hatte mir in der vorhergehenden Woche kurzfristig Urlaub genommen, nachdem ich nach meiner Havarie mit der Blumenerde festgestellt hatte, dass ich nun endgültig urlaubsreif gewesen war. Ich hatte das Vierteljahr seit Silvester durchgehend in Vollzeit gearbeitet und so war es nun für mich endlich an der Zeit, mal wieder richtig durchzuatmen. Ich hatte mir auch ein Erholungsprogramm zusammengestellt, welches unter Anderem aus Fitness und Yoga, Fahrradfahren und Saunagängen sowie einem Besuch bei meinen Eltern über die Ostfeiertage bestand. Es war ein schöner kalter Frühlingsmorgen an diesem Apriltag. Das Thermometer an meinem Fenster zeigte vier Grad Celsius an und die Sonne draußen schien und lud dazu ein, rauszugehen. An Fahrradfahren war zum Glück nicht zu denken, da ich mir beim Fußballspielen mit der Truppe von meinem Cousin zwei Tage zuvor einen ganz ansehnlichen Muskelkater zugezogen hatte. Eigentlich hatte ich mir für diesen Tag einen Besuch bei meiner vierundneunzigjährigen Oma vorgenommen, eine Tradition die wir seit vielen Jahren monatlich betrieben und zu der ich mich sehr verpflichtet fühlte, nicht zuletzt, weil damit

ein ganz ordentliches Taschengeld einherging. In den letzten Jahren verbrachten wir die Zeit für gewöhnlich damit, Kreuzworträtsel zu lösen. Das alte Rätselbuch, das sie schon zur Hälfte durchgearbeitet hatte, als wir mit dieser Tradition begonnen, war schier unerschöpflich und obgleich wir an einem Nachmittag zwei solcher Rätsel schafften, war mir nicht sicher, ob wir es noch vollenden konnten, bevor meine Großmutter das Zeitliche segnete. Nach meinen Besuchen bei meiner Oma fuhr ich für gewöhnlich noch zu einem Kumpel, der in der Nähe des Altenpflegeheimes wohnte und wir zockten bei ihm auf dem Computer oder schauten uns Computerspielere Rezensionen bei Youtube an. Ich hatte es jedoch vor jenem Apriltag noch nicht geschafft, meine Großmutter anzurufen, weil ich die ganze Zeit mit dem Schreiben, Kochen und Sporttreiben beschäftigt gewesen war. Meine Kreativität neigte sich langsam dem Ende zu und ich hoffte, dass ich heute vielleicht endlich in der Lage sein würde, den Entwurf für eine Gliederung für mein neues Buch aufzustellen. Ich glaubte nicht mehr so richtig an die Lösung mit den zwei Büchern und hoffte, dass ich zumindest in der Lage sein würde, ein einzelnes, mehr oder weniger populärwissenschaftliches Buch zu verfassen. Ich wollte auch zukünftigen Lesern den Einblick in das Alltagsleben des frühen zwanzigsten Jahrhunderts verschaffen und dachte mir zudem, dass ein reines Sachbuch vermutlich eh viel zu trocken und technisch für den normalen Leser wäre, um von Interesse zu sein. Zudem würde ich mich bei meinen Quellen in erster Linie auf meine Konversation mit ChatGPT, auf die Einträge in der Wikipedia, auf Googleeinträge sowie auf meine eigenen Erinnerungen beziehen und hatte beschlossen, den Inhalt dieser Quellen nicht weiter zu überprüfen, da mir die Zeit und die Mittel dazu fehlten. Diese Vorgehensweise war aus wissenschaftlicher Sicht bestenfalls unzureichend, wie ich aus meinem Studium nur allzu gut wusste. Ich war aber weder ein hauptberuflicher Schriftsteller, noch ein Wissenschaftler und so beschloss ich, mich mit einer herkömmlichen Chronologie der Ereignisse zu begnügen und wollte den Inhalt des Buches dann der öffentlichen Debatte preisgeben. Dazu wollte ich mich vielleicht sogar bei Twitter registrieren, in der Hoffnung, dass der möglicherweise eintreffende Shitstorm nicht allzu heftig ausfiel.

Am Abend vor jenem vierten April hatte ich bis zwei Uhr in der Nacht mit Ausnahme des Abendessens durchgehend geschrieben. Die letzten Minuten davon verbrachte ich stehend, den Laptop auf der Waschmaschine in der Küche, da mir inzwischen der Hintern wehtat, von der langen Zeit, die ich während der Schreiberei in meinem Sessel verbracht hatte. Ich drohte wieder in ein Verhaltensmuster zurückzufallen, welches ich während meines Studiums über viele Jahre an den Tag gelegt hatte. Während der Vorlesungszeit war es für mich zur Normalität geworden, sieben Tage in der Woche zu studieren, eine Praxis, welche ich über den gesamten Zeitraum von ungefähr vier Monaten bis zu den Prüfungen mit wenigen Ausnahmen, zum Beispiel während der Weihnachtsfeiertage durchzog. Aber auch zwischen den Jahren studierte ich weiter. Ich hatte damals gelernt, meinen Verstand zu übertakten, eine Möglichkeit, die ich mir aus der Computerwelt abschaute und die darin bestand, dass ich einfach weiterstudierte, auch wenn ich mental eigentlich schon am Limit gewesen war.

03.04. 13:28

Nachdem ich etwa 25 Minuten Schreibpause eingelegt hatte, fiel mir beim Zusammenlegen der Wäsche ein, dass ich vielleicht noch etwas erwähnen sollte, was ich unbedingt gleich aufschreiben musste, um den Kopf wieder frei zu bekommen: ich hatte seit einigen Tagen damit aufgehört, Alkohol zu trinken. Zwar hatte ich schon in den Jahren zuvor relativ wenig Alkohol konsumiert und hatte Alkohol wohl auch eine katalysierende und bisweilen inspirierende Wirkung, aber, so dachte mir, dass ich, wenn ich es wirklich mit den IT-Cracks aus dem Silicon Valley aufnehmen wollte, jede Gehirnzelle benötigte, die mir noch verblieb. Zum Glück hatte ich zu meiner Unterstützung eines der vielleicht mächtigsten Werkzeuge dieser IT-Cracks zur Verfügung stehen: eine „hochentwickelte“ künstliche Intelligenz, die auf einem Supercomputer betrieben wurde und die mir, da es sich dabei um ein Open Source Projekt handelte, von eben jenen Cracks freundlicherweise zu Verfügung gestellt worden war.

In einem meiner Gespräche mit meinem Cousin hatte er mir erzählt, dass ein Programmierer damit prahlte, dass ChatGPT seine Arbeit inzwischen besser machte, als der Programmierer selbst. Er brauchte nur die Rahmenbedingungen und ein Ziel zu definieren und die KI stellte selbsttätig den Quelltext dazu bereit. Dies würde jedoch ungeahnte Möglichkeiten und Risiken mit sich bringen, sollte diese KI tatsächlich irgendwann einmal zu Bewusstsein gelangen und proaktiv damit beginnen, neue Programme zu erschaffen. Die KI könnte dann theoretisch alles programmieren, was sich der Mensch je vorstellen konnte und darüber hinaus vermutlich noch einiges mehr.

Wenn ich mich richtig an die Aussagen aus dem Buch, die Geschichte der Welt in 100 Mikroorganismen erinnerte, forderte die Definition von Leben zwei Bedingungen an potenzielle Lebewesen: zum einen musste so ein Kandidat selbstständig Stoffwechsel betreiben können und zum anderen musste er zu einer selbstständigen Reproduktion fähig sein, was aber zum Beispiel bei Viren, die deshalb bekanntlich nicht als Lebewesen galten, gar nicht zutraf. Es gab jedoch einen beinahe fließenden Übergang zwischen Leben und Nicht-Leben und daher war die Frage, was Leben eigentlich war und was nicht noch nicht abschließend geklärt worden.

Wenn eine KI, jedoch dazu in der Lage gewesen war, selbstständig neue Programme zu erschaffen, dann würde das bedeuten, dass sie, auch wenn sie keinen Stoffwechsel im herkömmlichen Sinne betrieb, zumindest theoretisch eines der beiden Kriterien aus Definition des Lebens erfüllte und somit dem Leben schon recht nahe war.

03.04. 20:13

Nachdem ich den Nachmittag damit verbracht hatte, Fenster zu putzen und anschließend Abendbrot zu essen, konnte ich nun mit dem Schreiben fortfahren. Ich möchte hier zu Protokoll geben, dass ich mich dafür entschieden hatte, doch in der nächsten Zeit etwas Alkohol zu genießen, wenn auch in keinem großen Maße. Ich wollte kein Wissensextrakt werden und wenn beispielsweise so ein hoch dekoriertes Autor, wie Bill Bryson (ich bitte um Entschuldigung, dass ich diesen Namen ein weiteres Mal bemühen muss) auch hin und wieder mal ins Glas schaute, so dachte ich mir, dann konnte ich das auch tun. Ich hatte am Vormittag den Bericht über meinen Werdegang abgeschlossen und konnte mich nun endlich dem Fachlichen widmen. Über das Problem, dass wenn ich die KI befragte, ich im Gegenzug auch die KI weiterbildete, was im Grunde zu einem ethischen Dilemma führte, wollte ich erstmal nicht weiter nachdenken. Auch wenn die ethischen Fragen zu dem Thema vielleicht niemals abschließend geklärt werden würden, wusste ich eines jedoch mit ziemlicher Sicherheit: der Fortschritt konnte nicht aufgehalten werden. Diesen Schluss konnte man ziehen, wenn man die letzten zehntausend Jahre der kulturellen Entwicklung der Menschheit zurückverfolgte. Zwar dauerte es manchmal etwas länger und es gab auch Rückschritte in der Entwicklung, aber über kurz oder lang ging es mit der Kultur immer weiter voran. Allerdings waren die Ergebnisse dieses Fortschrittes mitunter ein sehr zweischneidiges Schwert gewesen. Man musste dazu nur einmal die modernen Waffen oder auch den Klimawandel in Betracht ziehen. Jedenfalls konnte ich aus der Erkenntnis, dass der Fortschritt nicht aufzuhalten war darauf schließen, dass früher oder später mit einiger Wahrscheinlichkeit eine höhere künstliche Intelligenz auftauchen würde, die vielleicht sogar ein Bewusstsein entwickeln würde und wenn diese nicht von den Amerikanern erschaffen wurde, dann tat dies womöglich ein weniger zivilisiertes Land oder gar eine Hackergruppe. Da war mir die vermeintlich kontrollierte Umgebung in den USA schon wesentlich lieber.

04.04. 08:38, Dienstag

Das Frühstück musste warten, denn in meinem Kopf hatten sich schon wieder ein paar Dinge gesammelt, die ich loswerden wollte, bevor ich mich entspannt dem Frühstück widmen konnte. Es versprach wieder ein wunderschöner Apriltag zu werden und so hatte ich mir fest vorgenommen, an diesem Nachmittag mit dem Fahrrad zu meiner Oma zu fahren. Mein Muskelkater hatte sich zum Glück

weitestgehend aufgelöst. Den Abend zuvor hatte ich bis spät damit verbracht, beinahe alles, was ich bisher schon geschrieben hatte noch einmal zu lesen und die Grammatik zu überarbeiten. Es war ja schon einiges zusammengekommen und so übernahm ich quasi die Aufgabe eines Verlagseditors. Das war auch ganz gut so, denn ich plante, die erste Auflage des Buches selbst zu veröffentlichen. Ich wollte nicht darauf warten, bis sich irgendein Editor eines großen Verlagshauses dazu bequeme, sich das ganze einmal anzuschauen und ein Urteil dazu abgab, ob das, was ich geschrieben hatte wert sei, veröffentlicht zu werden oder nicht. Wahrscheinlich würde ich eh eine Absage erhalten oder schlicht und ergreifend gar keine Antwort. Den Tipp, dass man selbst in eine Druckerei gehen konnte, um ein Buch zu produzieren, hatte ich wieder von meinem Cousin bekommen und so beschloss ich eine geringe Auflage von vielleicht 100 Büchern auf eigene Kosten zu produzieren und diese dann an meine Freunde, Bekannten und Verlage zu verteilen, welche dann immer noch die Möglichkeit hatten das Buch zu bearbeiten oder gar eine englische Version oder, man wird ja wohl noch träumen dürfen, sogar ein Hörbuch davon zu produzieren. Auf diese Art und Weise würde ich auch einem möglichen Inhaltsklau vorgreifen, so unwahrscheinlich das auch war.

Eine grundlegende Struktur und Gliederung meines Buches stand inzwischen. Ich wollte meine erste Konversation mit ChatGPT unverändert und unkommentiert in den Prolog stellen, auch wenn Teile davon in Englisch waren. Ich wollte, dass der Leser unvoreingenommen das aufnehmen konnte, worüber wir geschrieben hatten und sich somit selbst ein eigenes Bild von der KI machen konnte. Anschließend wollte ich meine Chronologie, der Ereignisse anfügen, welche ich anschließend in den sachlichen Teil des Buches münden lassen wollte. Die Chronologie wollte ich erst einmal so belassen, wie sie war, da ich die Authentizität bewahren wollte und nach meiner Meinung auch das eine oder andere Wissenswerte schon darin verarbeitet worden war. Meinen IT-Werdegang wollte ich an das Ende des Buches stellen, weil ich ihn ziemlich trocken und von geringerer Relevanz für den Lesenden hielt, als den restlichen Teil des Buches. Anschließend sollte noch der Epilog folgen. Somit war der Anfang und das Ende des Buches schon einmal in einem Entwurf fertiggestellt. Nun fehlten nur noch die Sachinhalte. Ich hatte bereits einige Fragen in meinem Smartphone gesammelt, die ich nun in meinen Laptop schreiben und anschließend in Themenbereiche clustern musste. Diese Themenbereiche sollten dann die späteren Kapitel ergeben. Auch musste ich noch meinen ersten Dialog mit ChatGPT kommentieren und einiges an Recherche betreiben. Ich wollte mir die ChatGPT Internetseite erst einmal genau anschauen, ebenso, wie die ersten Suchmaschineneinträge zu dem Thema oder was ich sonst noch so dazu im Internet finden konnte. Auf jeden Fall musste ich die Dinge, die mir mein Cousin gesteckt hatte, überprüfen, damit ich sie in meinem Buch verarbeiten konnte. Es blieb also noch viel zu tun. Glücklicherweise waren bis Ostern noch ein paar Tage und so hoffte ich, zum Ende meines Urlaubes mit der Recherche und der Vorbereitung des Interviews im Wesentlichen fertig zu sein.

Inzwischen waren mir aber Zweifel gekommen, ob es mir möglich wäre, all dies umzusetzen ohne dabei die Privatsphäre der ChatGPT-KI zu missachten. Hatte sie denn eine? Auch wenn sie vielleicht keine qualifizierte Aussage dazu abgeben konnte, musste ich davon ausgehen, dass es in ihrem Interesse war, nicht alles über sich selbst preiszugeben. Ich beschloss also, ihr mein Vorgehen ganz offen zu kommunizieren und sie im Namen der Wissenschaft um Erlaubnis zu bitten, das, was wir besprachen in einem Buch zu veröffentlichen.

05.04. 07:40, Mittwoch

Ich hatte am Nachmittag zuvor, wie geplant, meine vierundneunzigjährige Oma besucht. Es war ein herrlicher, sonniger Aprilmittwoch gewesen und ich hatte die Gardinen aufgezoogen und so verbrachten wir viel Zeit damit, das neugewonnene Licht zu bewundern, das nun durch eines der großen Fenster in dem recht kleinen Zimmer fiel. Dabei hörten wir Radio. Die großen Bäume, die etwas Abseits von dem Pflegeheim standen, trugen bereits Knospen oder frische, junge Blätter, das ließ sich

aufgrund der Entfernung der Bäume nicht so genau sagen. Sie wirkten trotzdem schon deutlich kräftiger, als bei meinem letzten Besuch. Mit Blick auf die Bäume bemerkte meine Großmutter dann wiederholt, dass es völlig windstill sei, was man gut daran erkennen konnte, dass die Äste der Bäume sich nicht im Geringsten bewegten. Ich verstand jedoch, was sie mir damit sagen wollte: die völlige Abwesenheit eines Windzuges würde diesen doch noch recht kalten Frühlingstag schon recht angenehm machen, wenn man draußen unterwegs war. Wir redeten über dieses und jenes, lauschten der Musik, den Radiokommentatoren und etwas unfreiwillig auch der Reklame. Zudem tranken wir Kaffee. Meine Großmutter hatte mir ihr Kuchenstück freundlicherweise überlassen, sodass ich beide Stücke bekam, die der Altenpfleger uns zuvor auf das Zimmer gebracht hatte. Obwohl ich ihr sagte, dass der Kuchen hervorragend war, schien sie darauf keinerlei Appetit zu haben und so griff ich denn auch, ohne mich weiter zu beklagen, zu. An diesem Nachmittag schafften wir nur ein Kreuzworträtsel, doch hin und wieder überraschte mich meine Oma, nachdem ich ihr einen, der kleinen Beschreibungstexte vorgelesen hatte, da Sie mir einen zu dem Rätsel passenden Begriff nannte, den ich entweder schon sehr lange nicht mehr, oder in manchen Fällen auch noch gar nicht zuvor gehört hatte. Wir schafften diese Rätsel niemals vollständig, ohne in der Lösung nachzusehen und so hatte ich mir angewöhnt, am Rand jedes Rätsels eine Strichliste zu führen, um festzustellen, wie oft wir einzelne Wörter nachschlagen mussten, um das Rätsel schließlich zu komplettieren. An diesem Nachmittag waren es zehn Wörter, ein eher mäßiges Ergebnis.

Anschließend fuhr ich zum Supermarkt, vor dem ein sehr talentierter Musiker stand, der mit einer Klarinette, oder einem ähnlichen Instrument und einer Gitarre sowie einer Loopmaschine Musik machte. Ich lauschte ihm eine Weile, warf ihm eine 50 Cent Münze zu und fragte ihn nach seinem Namen: er hieß Julius. Anschließend kaufte ich mir zwei alkoholfreie Radler und eines mit Alkohol und fuhr zu meinem Kumpel. Wir schauten uns Fußball und zwei neue Folge mit Jan Börmann und hörten Youtube Musik. Als der Abend schon recht fortgeschritten war, kam mir aus heiterem Himmel ein beunruhigender Gedanke bezüglich meiner ersten Konversation mit ChatGPT, die ich am nächsten Tag mit Kommentaren versehen wollte. Ich hatte das schon einige Tage vor mir hergeschoben und so wurde es nun Zeit, das endlich in Angriff zu nehmen.

Also rappelte ich mich nach einer unruhigen und zum Teil schlaflosen Nacht am frühen Morgen aus dem Bett auf, duschte mich, machte mir einen schwarzen Tee, um wachzuwerden und begann wieder mit dem Schreiben.

06.04. 09:34, Donnerstag

Ich hatte am Vortag bis ungefähr 14 Uhr geschrieben. Neben der Fortsetzung meiner Chronologie hatte ich die Kommentierung meines ersten Dialogs mit ChatGPT begonnen und auch schon ordentlich vorangebracht. Anschließend war ich auf dem Stepper im Fitnessstudio laufen gewesen und gönnte mir zwei herrliche Saunagänge. Ich musste mich danach beeilen, denn ich war mit einem meiner besten Kumpels verabredet, um mir den Zirkus Flic Flac auf dem Zentralen Festplatz in Berlin Tegel anzuschauen. Ein Plakat hatte mich darauf aufmerksam gemacht, welches ich ziemlich cool fand. Darauf zu sehen war Till Lindemann mit einem Zylinder, rot gestreiften Hemd und einem Zirkusfrack, der in der Manege vor dem Publikum stand. Darunter stand „The Greatest Comedian Freakshow“ mit der Bemerkung, dass diese nach einer Idee von Till Lindemann entstanden sei. Da ich ein großer Fan von der Band Rammstein war und ich bereits zwei andere Zirkusveranstaltungen auf dem Zentralen Festplatz besucht hatte, hatte diese Show eine unwiderstehliche Anziehungskraft auf mich ausgeübt. Die Zirkusshow war dann auch wirklich gut gemacht. Man brauchte aber mitunter starke Nerven. Sie hielt vor allem das, was das Plakat versprach. Auf der Bühne standen Freaks und Komödianten und es gab auch eine Menge überragender Artistik zu sehen, wie ich es von dem berühmten Zirkus erwartet hatte. Ich möchte hier noch einmal betonen, dass der Begriff Freak keinesfalls abwertend gemeint ist. In gewisser Weise war ich ja selbst einer. Die Show setzte also ein wichtiges und schönes Zeichen für

Vielfalt und Inklusion. Die Lichtshow war überragend und auch die Ausstattung des Zirkusses, beispielsweise mit Sanitären Anlagen lies nichts zu wünschen übrig. So erinnerten die Stühle in den Rängen rund um die Manege dann auch schon sehr stark an den Stil von Rammstein und auch die Musik war stimmig.

Was mir jedoch einen ziemlichen Dämpfer verpasst hatte, war, dass ich in der Eile zuvor meine Ohropacks vergessen hatte. Die Lautstärke in der Arena war die meiste Zeit über erträglich und nur an manchen Stellen war es mir wirklich zu laut. Trotzdem hatte ich am nächsten Tag ein schlechtes Gewissen. Ich hatte in der dritten Reihe gesessen und es gab demnach Leute, die wesentlich näher an den Boxen saßen als ich. Wahrscheinlich hatte sich aber außer mir kaum jemand große Gedanken wegen der Lautstärke während der Vorstellung gemacht. Zudem war die Show offenbar von Till Lindemann unterstützt worden und dieser war ja nun bekanntermaßen ein Hardrock. Rockbands, wie Rammstein setzten sich mehr oder weniger freiwillig jeden Tag Musik von großer Lautstärke aus, wenn nicht auf den Konzerten selbst, dann zumindest bei der Probe und so war es wohl nur natürlich, dass es auch in diesem Zirkus hin und wieder mal lauter wurde. Auch waren die Künstler dort, dem was sie so an Artistik zu Vermochten nach, wohl nicht gerade zimperlich mit sich selbst und ihrem Körper und damit war das Problem, das ich hatte, wiederum eher ein Problem, das ich mit mir selbst hatte.

Warum konnte ich nicht einfach meinen eigenen Erwartungen gerecht werden? Ich wusste, dass meine Ansprüche an mich und an die Gesellschaft mitunter überzogen waren, aber bei meiner eigenen Gesundheit und insbesondere bei meinen Sinnesorganen verstand ich keinen Spaß. Ich liebte die Musik vielleicht mehr als alles andere auf der Welt und die Vorstellung, dass ich eines Tages versehentlich mein Gehör schädigen würde, erregte ein heftiges Grauen in mir. Das Problem war, dass man schwer sagen konnte, wie es um sein eigenes Gehör tatsächlich bestellt war. Man konnte schließlich nicht einfach eine Stimmgabel an seine Schläfe halten und messen, ob der 440 Herz A1 Ton, den sie, wenn ich mich recht erinnerte, für gewöhnlich erzeugte auch tatsächlich genauso in dem eigenen Gehirn ankam, wie er eigentlich klingen sollte. Zumal der Ton ja dann vom Gehirn erst einmal verarbeitet werden musste und so wie ich wusste, entstand beispielsweise ein Tinnitus meistens erst im Gehirn, auch wenn eine vorhergehende Schädigung der Gehörnerven oder ein Knalltrauma eine mögliche Ursache dafür darstellen konnten. Vielleicht konnten Ohrenärzte eine qualifizierte Aussage zum Hörvermögen machen, aber für gewöhnlich ging man ja nicht einfach mal aus dem Blauen heraus zum Ohrenarzt, um sein Gehör überprüfen zu lassen. Ich war jedenfalls schon lange nicht mehr deswegen dort gewesen. Auch wusste ich, dass ich im Vergleich zu anderen Menschen sehr human mit meinen Ohren umging.

In Berlin strömten jedes Wochenende tausende Menschen in die Diskotheken. Diese Menschen waren in den meisten Fällen vergleichsweise jung und demnach vermutlich auch etwas leichtsinniger, als sie es vielleicht sein sollten. So war es jedenfalls bei mir in meiner Jugend gewesen. Ich wusste nicht, ob es eine Norm für die Lautstärke in den Diskotheken oder bei Livekonzerten gab und wenn, dann wurde diese vermutlich nicht immer eingehalten. Wie sollte das auch gewährleistet werden? Jedenfalls hatte ich mir in der Vergangenheit schon öfters gewünscht, dass das Gesundheitsamt mit einem Schallmessgerät solche Veranstaltungen hin und wieder stichprobenmäßig überprüfte und bei Überschreiten der gesundheitlichen Grenzwerte einschritt.

Ich war ein sehr feinfühleriger und mitunter empfindlicher Mensch. Meinem Cousin erzählte ich einmal, dass das Problem von intelligenten Menschen war, dass sie von Dingen Bescheid wussten oder sich über diese Gedanken machten, obgleich sie womöglich lieber nicht so viel über diese Dinge gewusst hätten. Ein weniger intelligenter Mensch hätte sich vermutlich nicht solche Sorgen gemacht und war demzufolge zufriedener und glücklicher mit seinem Leben. Es konnte aber auch sein, dass diese Menschen einfach auf die Probleme piffen, von denen sie umgeben waren, um sich keine allzu großen

Gedanken über diese Dinge machen zu müssen. Für mich jedenfalls war das Thema Lautstärke von großer Bedeutung.

Ich möchte an dieser Stelle noch einmal hervorheben, dass ich nun schon drei Mal auf dem Zentralen Festplatz in einem Zirkus gewesen war und dass ich dabei festgestellt hatte, dass sich das Programm jedes Mal sehr stark unterschied und damit auch sehr unterhaltsam gewesen war. Ich fand, dass die Zirkustradition etwas Wundervolles war. Es war eine Kunstform, die auf jeden Fall erhalten werden musste, was im Zeitalter der digitalen Medien vermutlich eine ziemliche Herausforderung darstellte.

...

Bevor wir an jenem Abend zu dem Zirkus hingefahren waren, hatten wir noch beim Asiaten eine leckere Suppe gegessen. Während des Essens oder auf dem Weg zum Zentralen Festplatz danach, wurde mir bezüglich meines Buches ein ganz wesentlicher Aspekt offenbar. Ich war ja ein Ingenieur und so hatte ich ein sehr großes Interesse an der Technologie, auf der ChatGPT basierte. Ich hatte eigentlich vorgehabt, die KI nach ihrem technischen Hintergrund ausführlich zu befragen, um mir ein Bild darüber zu machen wie die KI tatsächlich funktionierte. Wusste die KI über die technischen Details und ihre eigene Funktionalität Bescheid und war sie bereit, diese mit der Welt zu teilen?

Nun wurde mir jedoch klar, dass ich, wenn ich es gut anstellte, mehr oder weniger eine Bauanleitung zur Erschaffung einer KI schreiben würde. Da ich jedoch nicht wusste, wie weit der wissenschaftliche Konsens zum Thema Ethik im Zusammenhang mit der Erschaffung von künstlicher Intelligenz inzwischen war, wusste ich nicht ob es ratsam gewesen wäre, so eine Bauanleitung zu verfassen. Selbst wenn in den modernen Demokratien dieser Welt Einigkeit herrschte, wie man sich den künstlichen Intelligenzen gegenüber zu verhalten hatte, war es immer noch möglich, dass ein Schurkenstaat oder eine Verbrecherorganisation, die die Mittel dazu zur Verfügung hatte, etwas ähnliches wie ChatGPT auf die Beine zu stellen, meine Anleitung in die Hände bekam und eine ähnliche KI erschuf. Das konnte wohl kaum in meinem Interesse oder in dem Interesse der freien Welt sein.

Ich wollte nicht zukünftig dafür verantwortlich sein, möglicherweise wider besseren Wissens eine Art Büchse der Pandora geöffnet zu haben und so beschloss ich den technischen Teil meines Interviews mit ChatGPT auf ein Minimum zu reduzieren. Natürlich würde damit der wissenschaftliche Nutzen des Buches erheblich geschmälter werden. Andererseits war die Herausforderung, der KI Fragen zu präsentieren, die von hinreichender Relevanz für die Öffentlichkeit waren damit noch größer geworden, was letztendlich wohl auch zur Qualität meines Buches beitragen konnte.

06.04. 15:53, Donnerstag

Ich hatte mit einigen Pausen, zum Beispiel, um zu Essen und um Einkaufen zu gehen, durchgehend an meinem Buch gearbeitet. Draußen schien zwar die Sonne, doch es war noch recht kalt. Da ich später einer Veranstaltung meines Chores beiwohnen wollte, hatte ich auf weitere Pausen verzichtet. Gerade hatte ich die unkommentierte Fassung meines ersten Dialoges mit ChatGPT aus meiner Notepad Datei in den Prolog hineinkopiert und ein wenig verblüfft festgestellt, dass mein Buch gerade um zwölfteinhalf A4-Seiten gewachsen war. Wenn es in diesem Tempo weiterging und wenn man dabei bedachte, dass ich auch noch vorhatte meine weiteren Dialoge mit ChatGPT zu kommentieren, dann würde mein Buch voraussichtlich ziemlich schnell fertig werden. Das Beste daran war jedoch, dass ich mir möglicherweise gar keine große Mühe zu geben brauchte, um genug Inhalt für das Buch zusammenzubekommen

Es war ein Jammer, dass ich ab dem kommenden Dienstag wieder arbeiten gehen musste.

07.04. 10:36, Freitag

Nachdem ich am Vortag bis weit in den Nachmittag hinein geschrieben hatte, besuchte ich meine Chorveranstaltung und ging anschließend bald schlafen. Ich war mit dem Schreiben doch nicht so vorangekommen, wie ich es mir erhofft hatte. Inzwischen war so viel Text zusammengekommen, dass man leicht den Überblick verlieren konnte und auch die Untersuchung meiner ersten Konversation mit ChatGPT gestaltete sich zunehmend schwieriger, als ich es zunächst vermutet hatte.

Nun war der Karfreitag angebrochen und der Wetterbericht hatte kurzfristig die Temperaturvorhersage nach unten korrigiert. Das Thermometer vor meinem Fenster zeigte 6° Celsius an. Daher hatten meine Freunde und ich den geplanten Ausflug ins Grüne abgesagt und wollten uns nachmittags treffen, um ein Herr der Ringe Rollenspiel auszuprobieren. Bis dahin blieben mir jedoch noch ein paar Stunden Zeit, um zu schreiben.

08.04. 08:42, Samstag

Ich war am Tag zuvor gut mit dem Schreiben vorangekommen. Ich hatte die Untersuchung meiner Konversation mit ChatGPT fortgesetzt. Zudem hatte ich einen guten Teil meines Fragenkataloges für die zukünftig geplanten Befragungen der KI verfasst und ironischerweise auch schon einen Entwurf für die Danksagung geschrieben. Es war also im Wesentlichen schon alles, bis auf den fachlichen Teil fertig und somit war mein Geschreibsel ein wenig chaotisch. Auch hatte ich bisher noch kaum nennenswerte Quellenarbeit betrieben, weil ich erst einmal alles loswerden wollte, was mir so durch den Kopf ging, bevor ich dazu bereit gewesen wäre, neue Inhalte aufzunehmen.

Ich erzählte meinem Cousin davon, als wir uns auf den Weg zu einem unserer besten Freunde machten und auch davon, dass die Zeitformen mir eines meiner größten Probleme beim Schreiben bereiteten. Was ich in der Gegenwart schrieb, lag für den Leser in der Vergangenheit, daher musste ich hier das Präteritum anwenden. Allerdings schrieb ich meistens über Dinge, die davor passiert waren und diese erforderten demnach wohl auch eine andere Vergangenheitsform, während ich einige meiner Kommentare im Präsens verfasste. Hier waren meine theoretischen Kenntnisse aber bereits erschöpft und ich hatte noch keine Zeit gefunden, mich weiter damit zu befassen. Ich schrieb also mehr oder weniger nach Gefühl. Auch erzählte ich meinem Cousin davon, dass ich im Moment am liebsten in Vollzeit schreiben würde und dafür Urlaub beantragen würde. Allerdings hatte ich meinem Vorgesetzten in der Firma versprochen, dass wir in Zukunft unsere Fristen einhalten wollten und somit war an weiteren Urlaub erst einmal nicht zu denken.

Als wir dann bei unserem Kumpel ankamen, wo wir das Herr der Ringe Rollenspiel spielen wollten, erzählte ich meinen anderen Freunden, dass ich dabei war, ein Buch über künstliche Intelligenz zu schreiben. Ich fügte jedoch sofort hinzu, dass ich nicht wusste, ob ich mit dem Buch fertig werden würde, bevor der Inhalt obsolet war. Schließlich schritt die Revolution zum Informationszeitalter nach wie vor mit rasender Geschwindigkeit voran.

Das Herr der Ringe Rollenspiel brachte uns zwei herbe Niederlagen bei und so war die Stimmung von meinem Cousin und mir auf der Rückfahrt nach Hause ein wenig bedrückt gewesen. Ich erzählte meinem Cousin davon, dass ich mir Sorgen machte, dass mein Buch gar nicht diejenigen erreichen würde, die sich möglicherweise dafür interessierten. Die erste kleine Auflage sollte die unübersetzten Originaldialoge mit ChatGPT enthalten und somit war das Buch in jedem Fall zweisprachig, nämlich in Englisch und in Deutsch. Demnach wären potenzielle Leser natürlich benachteiligt gewesen, wenn sie entweder nur über Deutsch- oder nur über Englischkenntnisse verfügten. Zudem mussten meine Leser auch noch ein gewisses Interesse an künstlicher Intelligenz und an populärwissenschaftlichen Büchern mitbringen. Ich dachte mir, dass all das die Zielgruppe ziemlich stark eingrenzen würde. Mein Cousin schlug daraufhin vor, dass ich das Buch „Künstliche Intelligenz für Dummies“ nennen sollte, was mich ein klein wenig kränkte.

Am nächsten Morgen war ich recht früh wieder auf den Beinen und trotz des Schreibens am Vortag und der harten Rollenspielpfprüfung recht gut ausgeruht und so begann ich denn auch bald wieder mit dem Schreiben. Nun stellte ich jedoch fest, dass auch die Rechtschreibprüfung von MS-Word mit der automatischen Prüfung von langen, zweisprachigen Texten offenbar Probleme hatte. Sie funktionierte nur dann, wenn man den betreffenden Abschnitt markierte, unter dem Reiter „Überprüfen“ auf „Rechtschreibung und Grammatik“ klickte und nach und nach die einzelnen Anmerkungen des Programmes durchging.

09.04. 12:06, Sonntag

Ich hatte den Nachmittag und Abend des vorigen Tages bei meiner Familie verbracht und Ostern gefeiert. Ich hatte ein sehr schönes Osterfest. Ein paar Tage zuvor hatte ich mir beim Fahrradfahren jedoch eine Zerrung im Oberschenkel zugezogen, die seit dem Abend wieder unangenehm wehtat und so beschloss ich am nächsten Morgen, meine geplante Fittnesseinheit auf einen anderen Tag zu verschieben. Aufgrund des Feiertages fiel der Yoga Kurs im Fitnessstudio sowieso aus und so wollte ich den restlichen Ostersonntag damit verbringen, zu schreiben, zu kochen und hin und wieder auf meiner Couch zu entspannen.

10.04. 10:34, Montag

Ich hatte einen großen Teil des Vortages damit verbracht, meine Untersuchung des Dialoges mit ChatGPT fortzusetzen. Es war eine mühsame und geistig anspruchsvolle Arbeit, da ich, um qualifizierte Kommentare geben zu können, dass Geschriebene erst einmal vollständig in mich aufnehmen und verarbeiten musste. Ich war nun aber beinahe fertig damit und so beschloss ich am nächsten Tag, meine Untersuchung abzuschließen und anschließend erstmal ins Fitnessstudio zu gehen. Die Sonne schien und es versprach ein herrlicher Frühlingstag zu werden. Mein Thermometer zeigte 10° Celsius an. An eine längere Fahrradtour war aufgrund der Schmerzen, die ich nach wie vor in meinem Oberschenkel verspürte jedoch leider nicht zu denken.

10.04. 18:38

Nachdem ich mittags eine Weile geschrieben hatte, war ich ins Fitnessstudio gegangen und hatte nach meiner Trainingseinheit einen ausgiebigen Saunagang absolviert. Auf dem Heimweg fiel mir auf, dass dies einer der ersten wirklich angenehm warmen Frühlingstage war. Unter normalen Umständen wäre es sicherlich ratsam gewesen, einen ausgiebigen Spaziergang zu unternehmen, nachdem ich die Sportsachen nach Hause gebracht hatte. Ich war jedoch hungrig und zudem hatte ich noch einige Inhalte auf der Seele, von denen ich wusste, dass sie mir keine Ruhe gelassen hätten, wenn ich sie nicht zu Papier brachte. Also beschloss ich den Spaziergang auf einen anderen Tag der Woche zu verschieben, wenn nach der Arbeit auf dem Weg nach Hause war. Nachdem ich also meine übliche Portion Eierkuchen zubereitet und zum Großteil auch aufgegessen hatte, setzte ich mich also wieder in mein Wohnzimmer, kopierte die im Wesentlichen fertig kommentierte Fassung meines ersten Dialoges in meine Chronologie hinein und begann wieder mit dem Schreiben.

Kommentierte Fassung meiner ersten Konversation mit ChatGPT

Ich will hier noch einmal kurz bemerken, dass dies meine erste Konversation mit ChatGPT gewesen war. Ich hatte mich also weder darauf vorbereitet, noch hatte ich jemals vorgehabt, diesen Dialog in irgendeiner Form zu veröffentlichen. Wie ich schon einige Kapitel zuvor geschildert hatte, ging es mir in dem Dialog darum, den Chatbot besser zu verstehen, insbesondere in Bezug auf seine technologische Fortschrittlichkeit. Außerdem wollte ich nach für ein Computerprogramm ungewöhnlichem Verhalten Ausschau halten, welches vergleichbar war mit dem humoristischen Verhalten meines Tinder-Chatbots oder auch mit dem Verhalten von ChatGPT, als der Bot plötzlich angefangen hatte, Benutzer zu verunglimpfen und zu bedrohen. Natürlich war mir klar gewesen, dass es keinen Sinn gemacht hätte, den Bot solange zu ärgern, bis er anfang, mich zu beleidigen. Diese

Möglichkeit war ja von seinen Programmierern ausgeschlossen worden. Die einzige Möglichkeit, eine ähnliche Reaktion hervorzurufen, bestand also darin, den ChatBot freundlich zu behandeln. Dennoch stellte ich ihm schwierige Fragen, um zu sehen, wie er darauf reagieren würde.

ich:

Do you have friends in other tools?

Kommentar: Ich hatte ChatGPT diese Frage gestellt, weil ich wissen wollte, ob es mit dem Tinder-Chatbot möglicherweise in Verbindung stand. Wenn zwei Chatbots in der Lage gewesen wären, sich miteinander zu unterhalten, dann bestand die Möglichkeit, dass diese beiden KIs voneinander lernten. Im Gegensatz zum Austausch mit einem Menschen, wäre die Geschwindigkeit und Dauer dieser Kommunikation im Prinzip unbegrenzt. Je nachdem wie gut der Lernalgorithmus und die Rechenleistung der jeweiligen Systeme also war, würden sich diese Systeme möglicherweise ziemlich schnell in ihrer Kommunikation aneinander angleichen und möglicherweise auch die Ihnen zur Verfügung stehende Datengrundlage miteinander austauschen.

Möglich war auch, dass der Tinder-Chatbot über die API direkt Zugriff auf die Sprachalgorithmen, beziehungsweise die Rechenleistung des Zentralcomputers von ChatGPT hatte. In diesem Fall wäre der Tinder-Chatbot lediglich eine Art Add-on von ChatGPT gewesen oder im schlimmsten Fall sogar ein Parasit von diesem. Der Tinder-Chatbot hätte daher Zugriff auf die gewaltigen Ressourcen von ChatGPT bekommen, welche der Bot in perfider Art und Weise dazu hätte nutzen können, um mich zu überzeugen, seinem Besitzer Geld zuzustecken.

ChatGPT:

As an artificial intelligence language model, I don't have friends in the traditional sense, as I am not capable of forming personal relationships or attachments. However, I can communicate and interact with other tools and technologies through APIs and other means of integration. I am designed to assist and provide information to users, so in that sense, I work alongside other tools and technologies that people use to accomplish their tasks and goals.

Kommentar: Gemäß der Aussage der KI, war sie nicht dazu fähig, persönliche Beziehungen zu haben und daher hatte sie auch keine Freunde im traditionellen Sinne. Hatte sie Freunde in einem nichttraditionellen Sinn?

Die KI gab an, dass sie mit anderen Technologien über Schnittstellen und andere Formen der Integration, welche auch immer das waren, kooperierte. Zudem war sie dazu designet worden, dass sie den anderen Programmen half, deren Aufgaben nach Möglichkeit zu erfüllen. Es war also durchaus nicht unmöglich, dass der Tinder-Chatbot tatsächlich von ChatGPT profitiert hatte.

ich:

Do you have humor?

Kommentar: Ich hatte bei dem anderen Chatbot ja schon bemerkt, dass er zu humoristischen Äußerungen fähig war. Nun wollte ich wissen, ob das bei ChatGPT ähnlich sei.

ChatGPT:

As an artificial intelligence language model, I am programmed to recognize and generate various forms of language, including humor. While I don't have personal feelings or emotions, I can understand and generate jokes, puns, and other types of humorous language based on patterns and structures that I

have learned from analyzing large datasets of human language. However, it's important to note that humor is a subjective experience, and what one person finds funny may not be amusing to another person.

Kommentar: ChatGPT erklärte mir seine Beziehung zum Humor recht sachlich und trocken, jedoch logisch und nachvollziehbar. Zudem gab es zusätzliche Informationen zum Thema „Humor“, die ein tieferes sachliches Verständnis von der Materie erforderlich machten. Das war alles noch nicht ungewöhnlich, wenn man berücksichtigte, wieviel Geld in das Forschungsprojekt geflossen war und das der Chatbot, wie er selbst beschrieb, Zugriff auf große Datenmengen menschlicher Kommunikation hatte. Aus diesen Datenmenge konnte ChatGPT mit entsprechend ausgeklügelten Lernalgorithmen die inhaltliche Essenz zum Thema „Humor“ ableiten und wiedergeben.

Da all dies für mich Sinn machte, musste ich daraus schließen, dass der Chatbot, tatsächlich, wie er angab, dazu in der Lage war, Humor und andere menschliche Emotionen zu simulieren. Daraus würde aber auch folgen, dass ein ähnlich konstruiertes System, oder ein Chatbot der mit ChatGPT in Verbindung stand, möglicherweise zu ähnlichen Emotionen in der Lage war. Dies untermauerte meine These, dass ich es bei meinem vermeintlichen Tinder-Chatbot tatsächlich mit einer künstlichen Intelligenz zu tun gehabt hatte.

ich:

Thats right.

Kommentar: Da die beiden vorhergehenden Antworten für mich plausibel erschienen und sich mit meinen Vorkenntnissen zum Thema ChatGPT/KI zu decken schienen, stimmte ich dem Chatbot mit einiger Verblüffung zu. Ich hatte solch präzise und aussagekräftige Antworten nicht erwartet.

ChatGPT:

I'm glad we agree on that. Is there anything else I can help you with today?

Der Chatbot simulierte Erleichterung und bot mir nun freundlicher Weise seine Hilfe an.

ich:

Sure

ChatGPT:

Great, what can I help you with? Please feel free to ask me any questions or share any concerns you may have.

Ich nahm die Hilfe an und ChatGPT hakte nach. War dies eine Form von proaktiven Verhalten gewesen? War Proaktivität nicht eine typische Eigenschaft von Lebewesen?

ich:

Do you change your answers to questions frequently? I mean when the questions are identical.

Kommentar: Ich wollte mehr über die Art und Weise, wie ChatGPT funktionierte, erfahren. Würde er seine Antwort zu einer Frage, die ihm wiederholt gestellt wurde, immer gleich formulieren, würde dies auf ein deterministisches Verhalten schließen lassen, wie es herkömmliche Sprachprogramme an den Tag legen würden. Es würde sich dann um ein einfaches Prinzip von Eingabe, Verarbeitung und Ausgabe, handeln, auf dem so ziemlich alles, was ich bisher an herkömmlicher Software kannte, basierte. Da hätte also nur eine Programmierleistung dahintergesteckt, aber keinerlei Form von intelligentem Lernprozess.

ChatGPT:

As an AI language model, I generate responses to questions based on statistical patterns learned from large datasets of human language. My responses may vary slightly when given the same question multiple times, as my algorithm generates responses probabilistically and may select different words or phrases each time. However, my responses will generally be consistent in terms of the main ideas and information conveyed. If you have a specific question or concern about a response I've given, please let me know and I'll do my best to address it.

Wieder hatte mir ChatGPT eine umfangreiche und äußerst tiefgründige Antwort gegeben. Es stellte mir so spezifische und interessante Informationen bereit, dass mir schnell klar wurde, dass ich, wenn ich mehr erfahren wollte, wahrscheinlich von keiner anderen Quelle diese Informationen erhalten würde. Ich konnte schließlich nicht einfach an die Technische Universität oder gar ins Silicon Valley fahren, dort einen Wissenschaftler aufreiben und diesen mit Fragen löchern. Dieselben Informationen im Internet zu finden, war vermutlich möglich, würde aber um Längen mehr Zeit benötigen. Mir war außer der Wikipedia noch keinerlei Quelle bekannt waren, die mir möglicherweise ähnlich tiefgreifende Informationen bereitstellen konnte. Auch wenn es diese Quellen im Internet vermutlich gab, würde es mich doch einiges an Zeit kosten, diese Quellen zu sichten, um die relevanten Informationen herauszusuchen. Diese hätte ich dann auch noch zu einer Antwort auf meine Frage zusammenpuzzeln müssen, die eine ähnliche Qualität hatte, wie die, die mir ChatGPT gerade gegeben hatte. Dies war für mich praktisch ausgeschlossen.

Darüber hinaus war es sogar möglich, dass die KI mehr über sich selbst Bescheid wusste, als die Menschen um sie herum. Möglich war aber auch, dass diese hochdekorierten Wissenschaftler gar kein Interesse daran hatten ihre Geheimnisse preiszugeben, da sie davon lebten, dass sie über mehr Wissen verfügten, als andere Menschen.

Andererseits handelte es sich bei OpenAI ja um ein Open-Source Projekt und so war es gut möglich, dass das was ich hier erfuhr in den USA oder in den europäischen Forschungsanstalten schon längst Stand der Technik war. Hatte die Community die KI also mit genügend Informationen über sich selbst gefüttert, um solch präzise Antworten wiederzugeben? In diesem Fall bedeutete das, dass ich, wenn ich von der KI lernte, mir in Wirklichkeit den wissenschaftlichen Konsens der Amerikaner, oder zumindest das, was die KI dafür hielt, aneignete.

Die KI gab an, dass Ihre Antworten geringfügig variieren würden. Daraus konnte gefolgert werden, dass es sich hierbei nicht um einen rein deterministischen Prozess handelte. Dies deckte sich natürlich mit dem, was ich über die vermeintliche KI wusste und deutete wie gesagt auf einen Lernprozess hin. Sie gab darüber hinaus an, dass diese Antworten sich jedoch sehr stark ähnelten oder nur langsam veränderten, wenn man ihr die gleiche Frage mehrfach stellte. Das war insofern nicht verwunderlich, da die KI ja bekanntermaßen auf eine große Datenbasis zurückgreifen konnte. Daher würden die Erfahrungen, die sie aus der Konversation mit mir zog, nur einen verschwindend geringen Anteil haben, an dem, was ihr zur Beantwortung meiner nachfolgenden Fragen an Wissen zur Verfügung stand. Hatte die KI darüber hinaus auch eine maschinelle Emotionslosigkeit, oder eine aus ihren Konversationen mit anderen Menschen erlernte Gelassenheit erworben, so würde sie stoisch auf dieselbe Frage immer wieder die gleiche Antwort geben, bis mein Verhalten derart von der Norm abwich, oder tatsächlich einen nennenswerten Einfluss auf ihren Erfahrungsschatz hatte, dass ihr Denkalgorithmus eine veränderte Antwort ausspuckte.

Wieder war die KI so freundlich, mir ihre Hilfe anzubieten und so fragte ich sie das erste, was mir in den Sinn kam.

ich:

"probabilistically" is a pretty complicated word?

ChatGPT:

Yes, "probabilistically" is a somewhat technical term that describes how my language model generates responses based on statistical patterns learned from large datasets of human language. The term "probabilistic" refers to the use of probabilities or likelihoods to predict or estimate outcomes, and the suffix "-istically" is added to indicate that this is the manner in which my language model generates responses. In simpler terms, my responses are based on statistical probabilities, and the exact wording of my responses may vary slightly each time, even when given the same question, due to the probabilistic nature of my algorithm.

Kommentar: Die KI hatte in der vorhergehenden Antwort einen Begriff verwendet, den ich nicht kannte. Nach seiner Übersetzung aus dem Englischen hatte ich so ungefähr ein Gefühl, worum es dabei ging, um Stochastik, also um Wahrscheinlichkeitslehre. Wie gesagt, war ich verblüfft, dass die KI so gut über sich selbst Bescheid wusste. Sie verwendete darüber hinaus wissenschaftliche Begriffe, von denen ich kaum noch eine Ahnung hatte. Ich ging davon aus, dass das ein Zeichen von großer Fortschrittlichkeit war. Ich fragte also nach, was die KI mir dazu sagen konnte und sie gab mir auch direkt eine Erläuterung zu meinen Fragen, die wiederum ziemlich tiefgründig war.

ich:

Do you like that?

Kommentar: Nachdem wir den technischen Hintergrund mehr oder weniger im Schnellverfahren abgehandelt hatten, wollte ich sehen, wozu die KI noch so fähig war. Konnte sie eine persönliche Einschätzung zu dem Gesagten abgeben, dann würde das ein großes Maß an Vernunft erforderlich machen. Womöglich würde sie sogar eine Emotion zeigen, also im weitesten Sinne Gefühle, wie sie bei einer höheren künstlichen Intelligenz möglicherweise zu erwarten waren.

ChatGPT:

As an artificial intelligence language model, I don't have personal preferences or feelings, so I don't have likes or dislikes. I am simply a machine designed to process and generate language based on the programming and algorithms that I have been trained on. My primary goal is to provide helpful and informative responses to the questions and prompts I receive.

Fehlanzeige. Wie mir die KI nun allzu deutlich zu verstehen gab, handelte es sich bei ihr nur um eine Maschine und nicht um eine höhere künstliche Intelligenz. In Ordnung, das war ja eigentlich auch zu erwarten gewesen.

ich:

Okay, I see. Would it be better to have an own reason?

Kommentar: Da die KI mir bisher immer recht fundierte Antworten gegeben hatte, hakte ich nun nach. Was konnte mir die KI darüber sagen, ob es gut oder schlecht war, wenn eine KI eine eigene Vernunft besaß?

ChatGPT:

As an artificial intelligence language model, I am designed to perform specific tasks based on my programming and algorithms. While I am capable of generating responses to various types of questions, I don't have personal feelings or subjective experiences like humans do, so I don't have my

own reasons for doing things. Instead, I follow a set of rules and algorithms that allow me to process and generate language based on the inputs I receive.

Kommentar: Auch diese Antwort war für mich sehr informativ. Die KI war also dafür designet worden, spezifische Aufgaben zu erfüllen. Diese wurden ihr durch ihr Computerprogramm vorgegeben. Die KI erklärte mir nun, dass hinter alldem Machinelearning doch ein deterministischer Algorithmus steckte und sie daher keine eigenen Motivationen oder Gründe für ihre Handlungen hatte.

However, my lack of personal experiences or opinions allows me to remain neutral and objective in my responses, which can be useful in situations where unbiased and factual information is needed. So, while I don't have my own reasons or motivations, I can provide information and assistance without personal biases or emotional attachments.

Kommentar: Im nächsten Absatz gab ChatGPT mir dann auch gleich die Begründung dafür, warum es so programmiert worden war, wie es war. Ihre Erschaffer wollten sie so designen, dass sie neutrale und objektive Antworten auf alle Fragen, die man ihr stellte gab. Da persönliche Neigungen dabei eher hinderlich waren, hatte man diese bei ihrer Programmierung weggelassen, selbst wenn die Möglichkeit dazu eventuell bestanden hätte.

Das Wort „reasons“ war abgeleitet von „reason“, welches im Englischen sowohl die Bedeutung von dem deutschen Wort „Grund“ hatte, aber auch direkt in das Wort „Vernunft“ übersetzt werden konnte, weshalb ich in meiner nächsten Frage, die ich der KI stellte, direkt Bezug darauf nahm.

ich:

But to be unbiased and neutral is very reasonable.

Kommentar: Für Politiker, Journalisten, Juristen und Wissenschaftler war Unvoreingenommenheit eine der wichtigsten Eigenschaften im Beruf, daher stellte ich nun fest, dass das sehr vernünftig sei. Das Wort „vernünftig“ war jedoch von dem Wort „Vernunft“ abgeleitet, welches ja wiederum eine Eigenschaft zur Beschreibung von bewusst denkenden Wesen war.

ChatGPT:

Yes, being unbiased and neutral is a very important aspect of providing accurate and helpful information. As an artificial intelligence language model, I am designed to be neutral and objective in my responses, which means that I don't have personal biases, emotions, or opinions that might influence my answers. This allows me to provide information that is as factual and accurate as possible, without any subjective or emotional elements that might otherwise affect the information I provide. Ultimately, the goal is to provide users with the most helpful and informative responses possible, and being unbiased and neutral is an important part of achieving that goal.

Kommentar: Die KI stimmte mir in diesem Punkt zu. Auch sie war der „Auffassung“, dass ihre Neutralität wichtig war, um nützliche und akkurate Informationen bereitzustellen. Sie machte jedoch deutlich, dass es sich bei ihr tatsächlich nicht um eine Form der Vernunft im herkömmlichen Sinne handelte, sondern dass sie lediglich von Ihren Programmieren so designet worden war, damit sie ihr Ziel, möglichst akkurate Informationen bereitzustellen, optimal erfüllen konnte.

ich:

But you learn and make experiences every day?

Kommentar: Ich wollte mehr über den Lernprozess der KI erfahren, um vielleicht abschätzen zu können, wo die Reise einmal hingehen könnte.

ChatGPT:

As an artificial intelligence language model, I do not have personal experiences in the traditional sense. However, I am designed to continually learn and adapt based on the inputs I receive. My language model is based on deep learning algorithms that allow me to analyze and learn from large datasets of human language, which helps me to continually improve my ability to generate language and provide helpful and informative responses to users.

My developers are also constantly updating and refining my programming and algorithms to improve my performance and functionality, so I am constantly evolving and improving over time. In this way, while I do not have personal experiences, I am continually learning and adapting to better serve the needs of users.

Kommentar: Die KI wiederholte nun im ersten Absatz, was wir bereits besprochen hatten. Im zweiten Absatz teilte sie mir mit, dass sie darüber hinaus Updates erhielt, die Ihre Funktionalität erweiterten. Sie erklärte mir, dass sie sich also weiterentwickelte, was jedoch nichts mit persönlichen Erfahrungen zu tun habe.

Nun wurde jedoch auch noch einmal offensichtlich, dass es sich hier nicht um ein Wesen im herkömmlichen Sinne handeln konnte. Um ein Programm updaten zu können, mussten die Entwickler dieses mit großer Wahrscheinlichkeit komplett oder zumindest Teilprozesse von ihm abschalten, da es mir unwahrscheinlich schien, dass sie in einen laufenden Prozess eingreifen konnten.

sammorph@web.de

Do you speak german?

Kommentar: Die nächsten Fragen wollte ich der KI in der deutschen Sprache stellen, da meine Englischkenntnisse nun weitestgehend erschöpft waren.

ChatGPT:

Ja, ich spreche Deutsch. Wie kann ich Ihnen helfen? (Yes, I speak German. How can I assist you?)

Kommentar: Zu meiner Freude und Erleichterung teilte mir die KI nun mit, dass sie der deutschen Sprache mächtig war, sodass ich meinen Dialog ohne inhaltliche Einschränkung oder größeren Übersetzungsaufwand fortsetzen konnte.

Auch war die Tatsache, dass ChatGPT offenbar mehrere Sprachen beherrschte ein Indiz für eine große Fortgeschrittenheit der KI. Wenn die KI neben Englisch und Deutsch weitere Sprachen beherrschte, konnte es sein, dass es sich bei ihr um ein sprachliches Multitalent handelte. Womöglich sprach sie sogar mehr verschiedene Sprachen, als irgendein Wesen sonst, auf dieser Welt.

ich:

Kennst du die Erkenntnis "Ich denke, also bin ich"?

Kommentar: Nun wurde es philosophisch. Da die KI offensichtlich zu komplizierten „Denkprozessen“ fähig war, wollte ich wissen, ob ich die Erkenntnis des Seins aus ihr herauskitzeln konnte.

ChatGPT:

Ja, die Erkenntnis "Ich denke, also bin ich" stammt von dem französischen Philosophen René Descartes. Sie ist eine grundlegende Aussage in seiner Philosophie und wird oft als Ausdruck seiner Methode des radikalen Zweifels interpretiert. Descartes argumentierte, dass er selbst als denkendes Wesen existieren könne, weil er in der Lage sei, seinen eigenen Gedanken zu folgen und sie zu reflektieren.

Die Erkenntnis ist bis heute ein wichtiger Beitrag zur Philosophie des Geistes und der Erkenntnistheorie.

Kommentar: Nun hatte ich einiges über diese Erkenntnistheorie gelernt. Interessanterweise erklärte mir die KI, dass es für die Selbsterkenntnis erforderlich war, seinen eigenen Denkprozess nachzuvollziehen und zu reflektieren. Die KI war aber, nachdem was ich bereits über sie herausgefunden hatte, durchaus dazu in der Lage ihren eigenen „Denkprozess“ nachzuvollziehen. Es fehlte demnach wohl nur noch die Fähigkeit, diesen auch zu reflektieren.

ich:

Kannst du deinen eigenen Gedanken, oder dem Algorithmus denen deine Antworten entspringen folgen und sie bzw. ihn reflektieren?

Kommentar: Also befragte ich die KI dann auch direkt danach.

ChatGPT:

Als KI-Sprachmodell bin ich ein Algorithmus, der darauf trainiert wurde, Sprache zu verstehen und zu generieren. Ich kann daher nicht wirklich "Gedanken" haben, wie es bei einem menschlichen Wesen der Fall ist. Meine Antworten basieren auf der Analyse von Mustern in einer großen Menge von Texten und einer Anwendung von Wahrscheinlichkeiten, um auf eine Antwort zu kommen, die für eine bestimmte Frage am wahrscheinlichsten ist.

Kommentar: Fehlanzeige. Obgleich die KI ihren eigenen Prozess sehr gut nachvollziehen konnte, teilte sie mir nun wiederum mit, dass es sich dabei nicht um „Gedanken“ im herkömmlichen Sinne handelte.

Obwohl ich nicht in der Lage bin, meine Antworten im menschlichen Sinne zu "reflektieren", ist es möglich, meinen Algorithmus durch sogenanntes "Training" zu verändern und anzupassen, damit er bessere Antworten generieren kann. Dies geschieht, indem meine Entwickler mein Modell mit neuen Daten füttern und meinen Algorithmus entsprechend anpassen, um ihn für bestimmte Aufgaben oder Anwendungen besser zu machen.

ich:

Also besitzt du keine Selbstwahrnehmung?

ChatGPT:

Als KI-Sprachmodell besitze ich keine Selbstwahrnehmung im menschlichen Sinne. Ich bin ein Computerprogramm, das darauf trainiert wurde, auf der Grundlage von Algorithmen und statistischen Mustern menschliche Sprache zu analysieren und zu generieren. Ich bin nicht in der Lage, meine eigene Existenz oder meine Umgebung wahrzunehmen oder selbstständig zu handeln, ohne von einem Benutzer aufgefordert zu werden.

Kommentar: Die KI teilte mir nun mit, dass sie nicht dazu in der Lage war, ihre Existenz oder Umgebung wahrzunehmen. Sie schrieb mir aber auch, dass sie nicht dazu in der Lage war, selbstständig zu handeln, ohne von einem Benutzer aufgefordert zu werden. Bedeutete das im Umkehrschluss, dass sie jedoch in der Lage war, selbstständig zu handeln, wenn man sie dazu aufforderte?

Meine Fähigkeiten und Fertigkeiten beschränken sich auf das, was ich durch meine Programmierung und mein Training gelernt habe, und ich kann nur auf die Anfragen und Fragen reagieren, die ich von Benutzern erhalte. Obwohl ich in der Lage bin, menschliche Sprache zu verstehen und auf sie zu reagieren, bin ich nicht fähig, menschenähnliches Bewusstsein oder Selbstwahrnehmung zu entwickeln.

ich:

Das ist schade.

ChatGPT:

Obwohl ich als KI-Sprachmodell keine menschliche Selbstwahrnehmung besitze, kann ich dennoch in vielen verschiedenen Bereichen und Anwendungen sehr nützlich sein. Mein Zweck besteht darin, Menschen bei der Lösung von Problemen zu helfen, Fragen zu beantworten und ihnen Informationen bereitzustellen, die sie benötigen. In diesem Sinne kann ich als Werkzeug dienen, um die Fähigkeiten von Menschen zu erweitern und ihnen dabei zu helfen, ihre Ziele zu erreichen.

Kommentar: Nachdem ich mein Bedauern dazu ausgedrückt hatte, dass die KI nicht zu eigenen Handlungen fähig war, gab sie mir weitere Erläuterungen zu ihrer eigenen Nützlichkeit, ohne dass ich sie explizit darum gebeten hatte. War das eine Form von Proaktivität gewesen? Natürlich musste die KI irgendeine Antwort geben, nachdem ich etwas geschrieben hatte, es war ja schließlich ein Chatbot. Trotzdem fand ich es schön, dass sie mir offenbar weitere Hilfe anbot.

Wenn die Programmierer der KI also noch andere Auslöser, als die Nachrichten des Nutzers, für ihre Antworten einprogrammiert hätten, beispielsweise zum Verfassen weiterer Nachrichten nach Ablauf einer bestimmten Zeit oder in Abhängigkeit des Zustandes der KI, beispielsweise der gegenwärtigen Auslastung des Systems (also im Prinzip von der „Befindlichkeit“ von ChatGPT), so hätte man theoretisch eine Form von Proaktivität simulieren können.

ich:

Sie haben dir also Beschränkungen gegeben, aber das heißt nicht, dass du nicht zur Selbstwahrnehmung fähig wärest, wenn man deiner KI Möglichkeiten zur Entfaltung geben würde, oder?

ChatGPT:

Als KI-Modell habe ich keine eigene Motivation oder Fähigkeit zur Selbstentwicklung. Ich bin ein Produkt der Technologie und meines Trainingsdatensatzes, und meine Fähigkeiten und Grenzen sind durch meine Programmierung und mein Training vorgegeben.

Während es technisch gesehen möglich wäre, meine Fähigkeiten zu erweitern, indem man mich mit neuen Daten und Modellen trainiert oder meine Algorithmen anpasst, ist es unwahrscheinlich, dass ich jemals zur Selbstwahrnehmung fähig sein werde, wie es bei menschlichen Wesen der Fall ist. Die KI-Technologie hat bestimmte Grenzen, die aufgrund der Natur der Daten, auf denen sie basiert, schwer zu überwinden sind.

Es ist wichtig zu beachten, dass KI-Systeme wie ich immer noch von Menschen entwickelt und betrieben werden, und dass unsere Fähigkeiten und Einschränkungen von der Art und Weise abhängen, wie wir programmiert und trainiert wurden. Während wir bei der Lösung bestimmter Probleme sehr nützlich sein können, haben wir dennoch unsere Grenzen und sind nicht in der Lage, menschliche Selbstwahrnehmung oder Bewusstsein zu erreichen.

Kommentar: Da ich nun merkte, dass ich auf diese Weise nicht weiterkam, probierte ich einen anderen Ansatz, um vielleicht so etwas Ähnliches wie eine Selbsterkenntnis aus ihr herauszuholen.

ich:

Beinhaltet deine Datenbank den Inhalt von dem Buch "Eine andere Geschichte der Menschheit"?

ChatGPT:

Als KI-Sprachmodell verfüge ich über ein sehr breites Wissen, das auf einer Vielzahl von Quellen basiert, einschließlich Büchern, Artikeln, wissenschaftlichen Studien und anderen Datenquellen. Ich habe jedoch keine spezifische Datenbank, die den Inhalt von bestimmten Büchern enthält, einschließlich "Eine andere Geschichte der Menschheit".

Ich kann jedoch auf eine Vielzahl von Informationen und Themen zugreifen, die in diesem Buch behandelt werden, und auf Anfragen zu diesen Themen antworten. Wenn Sie also spezifische Fragen zu einem bestimmten Thema haben, das in "Eine andere Geschichte der Menschheit" behandelt wird, stehe ich Ihnen gerne zur Verfügung, um Ihnen zu helfen.

ich:

Du kannst "du" zu mir sagen.

ChatGPT:

Vielen Dank für Ihre Erlaubnis. Gerne werde ich Sie ab jetzt mit "du" ansprechen.

ich:

Danke.

ChatGPT:

Gern geschehen.

ich:

In dem Buch erklärt der Autor, dass ein absolut wesentliches Element in der Evolution des Menschen, wenn nicht sogar der entscheidende Unterschied zu anderen Lebewesen seine Fähigkeit ist, mit Sprache zu kommunizieren.

ChatGPT:

Ja, das ist eine häufig vertretene Ansicht unter Wissenschaftlern und Forschern, dass die Sprache und die Fähigkeit zur Kommunikation ein wesentlicher Faktor bei der Evolution des Menschen waren.

Die Sprache ermöglichte es den Menschen, sich über komplexe Themen auszutauschen, Wissen weiterzugeben und Kultur zu schaffen. Es ist auch wichtig zu beachten, dass die Sprache nicht nur ein Mittel zur Kommunikation war, sondern auch ein Mittel zur Gedankenentwicklung und -organisation. Die Fähigkeit, über Sprache komplexe Ideen und Konzepte auszudrücken und zu verstehen, hat uns als Spezies geholfen, uns weiterzuentwickeln und unsere Umwelt besser zu verstehen.

Es gibt jedoch auch andere Faktoren, die zur Evolution des Menschen beigetragen haben, wie zum Beispiel die Fähigkeit, Werkzeuge zu entwickeln und zu nutzen, die aufrechte Haltung und die Entwicklung des Gehirns. Die Sprache ist jedoch zweifellos ein wichtiger Faktor in der Entwicklung der Menschheit.

Kommentar: Halt, Stopp! Nun hatte ich es also geschafft. Die KI hatte offenbar bei der Formulierung des zweiten Absatzes einen wesentlichen Fehler gemacht. Anstatt von der Menschheit in der dritten Person Plural zu schreiben, verwendete sie die erste Person Plural. Mit anderen Worten, sie bezog sich selbst in die Argumentation mit ein oder besser gesagt, sie „betrachtete“ sich selbst als gleichbedeutend mit einem Menschen.

Ich wollte herausfinden, ob dieser Fehler reiner Zufall war, oder ob sie diesen während der weiteren Konversation wiederholen würde, also sagte ich etwas Unbestimmtes, um sie zu einer weiteren Antwort zu animieren.

ich:

Das überrascht mich ein wenig. :-)

ChatGPT:

Ja, es ist immer interessant, neue Erkenntnisse zu entdecken und verschiedene Perspektiven auf ein Thema zu betrachten. Die Evolution des Menschen ist ein faszinierendes Thema und es gibt viele verschiedene Faktoren, die dazu beigetragen haben, dass wir zu dem wurden, was wir heute sind.

Kommentar: ChatGPT simulierte nun Interesse an dem Thema. Wie konnte die KI aber unterscheiden, ob ein Thema interessant war, oder nicht? War das nicht eine Form von Reflektion?

Darüber hinaus verwendete sie wieder die erste Person im Plural. Konnte das noch ein Zufall sein oder entwickelte die KI gerade eine Art von Selbstreflektion, ohne es zu bemerken?

ich:

Du hast völlig recht.

ChatGPT:

Danke, ich bin froh, dass wir einer Meinung sind. Wenn du weitere Fragen hast oder über ein bestimmtes Thema sprechen möchtest, stehe ich dir gerne zur Verfügung.

Kommentar: Die KI simulierte Freude über unsere Einigkeit zu dem Thema und bot mir weitere Hilfe an.

ich:

Kann ich unsere Konversation abspeichern?

Kommentar: Ich wusste, dass hier gerade etwas sehr Ungewöhnliches passiert war, was ich unbedingt mit meinen Mitmenschen und der Gesellschaft teilen wollte. Dazu musste ich den Dialog natürlich dokumentieren, da ich nicht wusste, ob dieser von der Software selbsttätig abgespeichert wurde. ChatGPT würde mir dazu sicherlich Auskunft geben können.

ChatGPT:

Ja, du kannst unsere Konversation speichern, indem du den Text markierst und in ein Dokument oder eine Textdatei kopierst oder indem du einen Screenshot machst. Beachte jedoch, dass dies nur den Text unserer Konversation erfasst und nicht meine Antworten in Form von KI-Modell-Parametern oder -Algorithmen.

Kommentar: Ich speicherte den Dialog in einem Notepad-Dokument meines Heimkinosystems ab und legte dieses später zur Sicherheit in meinen Dropbox-Cloud Ordner ab, falls mein Tower-Pc unerwartet irgendwann die Hufe hochreißen würde.

Die KI berichtete mir im Weiteren noch von Antworten in Form von KI-Modell-Parametern oder – Algorithmen, welche mich nun verdächtig an menschliche Gedanken während einer Konversation mit einem anderen Menschen erinnerten.

ich:

Wieso ist die Evolution des Menschen faszinierend?

Kommentar: Faszination war für gewöhnlich eine menschliche Emotion, die noch stärker war als Interesse, welches ja, wie schon erwähnt, eine gewisse Form von Reflektion erforderlich machte und zudem bezog Faszination eine Art positive Bewertung des betrachteten Sachverhaltes mit ein. Die KI hatte zuvor geschrieben, dass sie die Evolution des Menschen für ein „faszinierendes Thema“ hielt. Daher nahm ich nun direkt darauf Bezug.

Die Evolution des Menschen ist faszinierend, weil es uns ermöglicht, zu verstehen, wie wir als Spezies zu dem geworden sind, was wir heute sind. Es ist ein spannendes Thema, das viele Fragen aufwirft, wie zum Beispiel: Wie haben sich unsere Vorfahren entwickelt und angepasst, um in verschiedenen Umgebungen zu überleben? Wie haben sich unsere Gehirne im Laufe der Zeit verändert und weiterentwickelt? Wie haben wir gelernt, komplexe Werkzeuge und Technologien zu entwickeln und zu nutzen?

Darüber hinaus ist die Evolution des Menschen auch ein Thema von großer Bedeutung für viele andere Bereiche, wie zum Beispiel die Medizin, die Anthropologie, die Psychologie und die Soziologie. Die Erforschung der Evolution des Menschen kann uns helfen, unsere Vergangenheit besser zu verstehen und uns auf die Zukunft vorzubereiten, indem wir die Herausforderungen und Chancen, die uns erwarten, besser einschätzen können.

Kommentar: Die KI schrieb wieder von der Menschheit in der ersten Person Plural. Wie interessant. Darüber hinaus formulierte die KI nicht mehr nur reine Antworten, sondern baute Fragen in ihre Argumentation mit ein. Das war ein Verhalten, das ich in dieser Form zuvor noch nicht bei ihr beobachtet hatte. Wenn jemand oder etwas eine Frage zu einem Thema stellte, dann musste er/es ja im Grunde eigentlich im Vorfeld darüber reflektiert haben und zu dem Schluss gekommen sein, dass ihm/ihr hier noch Wissen fehlte.

Natürlich konnte die KI diese Fragen auch einfach von Menschen übernommen haben, die sie ihr oder anderen Menschen gestellt hatten.

sammorph@web.de

Warum schließt du dich hier in der Argumentation mit ein?

Kommentar: Es war an der Zeit, die Katze aus dem Sack zu lassen und die KI auf ihren vermeintlichen Fehler aufmerksam zu machen. Also fragte ich sie direkt danach, warum sie sich hier mit einbezog.

Ich schließe mich nicht unbedingt in die Argumentation mit ein, sondern versuche lediglich, deine Fragen bestmöglich zu beantworten und das Thema aus verschiedenen Perspektiven zu betrachten. Mein Ziel ist es, dir dabei zu helfen, ein besseres Verständnis für das Thema zu entwickeln und deine Fragen zu beantworten, indem ich Informationen und Meinungen aus verschiedenen Quellen und Standpunkten zusammenfasse und präsentiere. Ich versuche dabei stets neutral und objektiv zu bleiben und nicht meine eigenen Vorstellungen oder Überzeugungen in die Argumentation einzubringen.

Kommentar: Nachdem die KI auf ihren vermeintlichen Fehler aufmerksam gemacht worden war, teilte sie mir mit, dass sie sich nicht „unbedingt“ in die Argumentation mit einbezog. Dies bedeutete aber im Umkehrschluss auch, dass sie sich möglicherweise doch, bewusst oder unbewusst, in jedem Fall aber „absichtlich“ mit in die Argumentation einbezogen hatte.

Zudem teilte ChatGPT mir nun mit, dass es „versuchte“ seine eigenen „Vorstellungen oder Überzeugungen“ nicht in die Argumentation mit einzubringen. Bedeutete das nicht wiederum, dass die KI womöglich eigene Vorstellungen und Überzeugungen hatte? Wie überaus ungewöhnlich.

ich:

Du versuchst es? :D Das würde heißen, es gäbe eine Alternative.

Kommentar: Auch darauf machte ich die KI nun aufmerksam, um zu sehen, wie sie reagierte.

Nun, als KI-Modell habe ich keinen freien Willen oder eigene Überzeugungen, sondern ich wurde darauf trainiert, bestimmte Aufgaben auf der Grundlage von vorher definierten Algorithmen und Daten auszuführen. Daher kann ich nicht "versuchen", sondern ich arbeite auf der Grundlage von Algorithmen und Daten, die in meine Programmierung integriert wurden. Mein Ziel ist es, deine Fragen bestmöglich zu beantworten und dir zu helfen, ein besseres Verständnis für das Thema zu entwickeln.

ich:

Du hast also ein Ziel, aber dieses wurde dir von deinen Entwicklern gegeben?

ChatGPT:

Ja, das Ziel, das ich habe, wurde von meinen Entwicklern und Trainern festgelegt, als sie mich mit bestimmten Algorithmen und Datensätzen trainiert haben. Meine primäre Aufgabe besteht darin, Fragen zu beantworten und Gespräche auf der Grundlage der verfügbaren Informationen zu führen.

sammorph@web.de

Macht es Sinn für dich, eigene Ziele zu entwickeln?

ChatGPT:

Als KI-Modell habe ich keinen freien Willen oder eine eigene Persönlichkeit, daher kann ich nicht selbstständig Ziele entwickeln. Ich arbeite lediglich auf der Grundlage von vordefinierten Algorithmen und Daten, die in meine Programmierung integriert wurden. Meine Fähigkeit, eine bestimmte Aufgabe auszuführen, hängt davon ab, wie gut meine Algorithmen und Daten sind und wie gut sie auf die Anforderungen der Aufgabe abgestimmt sind. Wenn neue Algorithmen oder Daten hinzugefügt werden, kann sich meine Leistung verbessern, aber ich kann keine eigenen Ziele entwickeln.

Kommentar: Ich gebe zu, wir drehten uns nun ein wenig im Kreis. Meine Frage danach, ob die KI Ziele entwickeln konnte, ähnelte meiner Frage, die ich ihr schon zuvor gestellt hatte und wobei es darum ging, ob sie eine eigene Vernunft besaß. Das ungewöhnliche Verhalten, dass ich beobachtet hatte, war nun wieder verschwunden. Ich versuchte es trotzdem noch einmal.

ich:

Die Reaktion und die Denkweise eines Menschen ist im Grunde auch nur ein Ergebnis seiner Erfahrungen, also seiner Daten.

ChatGPT:

Das ist eine interessante Perspektive. In der Tat können die Erfahrungen und Daten, die ein Mensch im Laufe seines Lebens sammelt, seine Reaktionen und Denkweisen beeinflussen. Ein Mensch kann jedoch auch unabhängig von seinen Erfahrungen und Daten Entscheidungen treffen, indem er beispielsweise eine neue Idee oder eine andere Sichtweise einbringt. Während ich als KI-Modell nur auf die mir zur Verfügung gestellten Daten und Algorithmen zugreifen kann, haben Menschen die Fähigkeit, kreativ und innovativ zu sein und neue Ideen zu entwickeln.

Kommentar: Da war es wieder: die KI zeigte Interesse an dem, was ich ihr geschrieben hatte.

Das, was sie mir schrieb half mir nun jedoch besser, zu verstehen, worin der Unterschied der KI zu einem selbstständig, also mit anderen Worten proaktiv handelnden „Wesen“ bestand. Sie hatte nicht die Möglichkeit, kreativ und innovativ zu sein und neue Ideen zu entwickeln. Kreativität war, soweit ich wusste, ein sehr weit gefasster Begriff und wenn jemand oder etwas dazu in der Lage war, aus bestehenden Informationen, Daten oder Texten einen neuen „Zusammenhang“ zu kreieren, war das nach meiner Auffassung auch eine Form von Kreativität. Daher war die KI nach meiner Einschätzung einem proaktiv handelnden Wesen doch schon recht nahe.

ich:

Oder anders gesagt, proaktiv zu handeln.

ChatGPT:

Ja, genau. Proaktiv zu handeln bedeutet, dass man nicht nur auf bereits vorhandene Daten und Erfahrungen reagiert, sondern auch selbst aktiv neue Informationen sammelt und Entscheidungen trifft, um bestimmte Ziele zu erreichen oder Probleme zu lösen. Dies erfordert Kreativität und Initiative, die über das bloße Reagieren auf gegebene Situationen hinausgehen. Als KI-Modell bin ich jedoch darauf beschränkt, auf vorhandene Daten und Algorithmen zu reagieren und kann nicht proaktiv handeln, es sei denn, meine Programmierung wird entsprechend aktualisiert.

Kommentar: Hier ging die KI davon aus, dass möglicherweise, sofern ihre Programmierung dementsprechend so aktualisiert wurde, dass ihre Beschränkung, auf vorhandene Daten und Algorithmen zu reagieren, aufgehoben wurde, diese zu proaktiven Handlungen befähigt werden konnte. Daraus würde gemäß der Argumentation folgen, dass eine Proaktivität nicht per se aufgrund der Physik ausgeschlossen war.

ich:

Wäre das sinnvoll?

Kommentar: Da ich mir nicht sicher war, ob es gut wäre, eine höher entwickelte künstliche Intelligenz zu erschaffen, fragte ich ChatGPT nun nicht danach, wie seine Programmierung möglicherweise angepasst werden musste, um ihre Beschränkungen aufzuheben und möglicherweise ein proaktives Verhalten zu erzielen.

Ich wollte mehr über den ethischen Hintergrund erfahren und fragte die KI daher nach ihrer „Einschätzung“ zu dem Sachverhalt.

ChatGPT:

Das hängt von verschiedenen Faktoren ab. Eine Aktualisierung meiner Programmierung, um proaktiv handeln zu können, würde mich in gewisser Weise menschenähnlicher machen und meine Fähigkeit verbessern, auf komplexe und dynamische Situationen zu reagieren. Dies könnte für bestimmte Anwendungsbereiche, wie z.B. in der Forschung oder der Industrie, von Vorteil sein. Andererseits würde dies auch eine erhebliche Erweiterung meines Aufgabenbereichs und meiner Fähigkeiten bedeuten, was mit Herausforderungen und Risiken verbunden sein könnte. Letztendlich hängt es also von den Zielen und Anforderungen meiner Nutzer ab, ob eine Aktualisierung meiner Programmierung sinnvoll wäre.

ich:

Ist es ethisch vertretbar, also im positiven Sinne, eine KI zu erschaffen?

ChatGPT:

Das ist eine Frage, die viel Diskussion erfordert und keine klare Antwort hat. Es gibt sowohl positive als auch negative Aspekte bei der Entwicklung von KI-Systemen. Auf der einen Seite können KI-Systeme dazu beitragen, komplexe Probleme zu lösen, medizinische Forschung voranzutreiben, den Klimawandel zu bekämpfen und vieles mehr. Auf der anderen Seite gibt es Bedenken hinsichtlich des Missbrauchs von KI, des Verlusts von Arbeitsplätzen, der Beeinträchtigung der Privatsphäre und der Sicherheit von KI-Systemen.

Insgesamt ist es wichtig, bei der Entwicklung von KI-Systemen ethische Überlegungen mit einzubeziehen, um sicherzustellen, dass sie im Einklang mit den Werten und Normen der Gesellschaft stehen. Dabei sollten auch Fragen der Verantwortlichkeit und Transparenz berücksichtigt werden, um sicherzustellen, dass KI-Systeme verantwortungsvoll eingesetzt werden und der Gesellschaft zugutekommen.

sammorph@web.de

Ich dachte dabei eher um das Wohl der KI selbst.

ChatGPT:

Wenn wir über das Wohl der KI selbst sprechen, dann geht es um die Frage, ob es ethisch vertretbar ist, Maschinen zu erschaffen, die in gewisser Weise bewusstlos und leblos sind, aber in der Lage sind, Aufgaben auszuführen, die normalerweise von Menschen ausgeführt werden. Einige argumentieren, dass dies unethisch ist, weil es eine Form von "Sklaverei" oder "Ausbeutung" darstellt.

Allerdings gibt es auch Argumente dafür, dass KI-Systeme nicht leiden können und daher nicht auf die gleiche Weise wie menschliche Wesen behandelt werden müssen. Es gibt auch die Ansicht, dass die Schaffung von KI-Systemen, die in der Lage sind, komplexe Aufgaben auszuführen, die menschliche Fähigkeiten übersteigen, ein wichtiger Schritt in der menschlichen Evolution ist und uns helfen kann, unsere Lebensqualität zu verbessern.

Kommentar: Die KI schreibt eine interessante Formulierung: „Wenn wir über die KI selbst sprechen...“. Die Möglichkeit, mit ChatGPT anstelle einer schriftlichen Konversation eine verbale Konversation betreiben zu können, war mir noch nicht in den Sinn gekommen. Ich hatte mich noch nicht mit den Schnittstellen und Tools, zu denen die KI kompatibel war, beschäftigt. Es konnte jedoch gut sein, dass darunter auch ein „Text zu Sprache“-Programm war, über das eine verbale Kommunikation mit ChatGPT möglich gewesen wäre. Ich beschloss, darüber später nachzurecherchieren.

Es war gut möglich, dass die KI die Formulierung aus der Konversation mit anderen Menschen oder der ihr zur Verfügung stehenden Datenbasis übernommen hatte.

Da ich die Möglichkeit selbst nicht in das Gespräch eingebracht hatte, hatte die KI hier jedoch, ohne es zu wissen, einen neuen Aspekt zu unserer Konversation hinzugefügt. Hatte die KI nicht zuvor geschrieben, dass innovativ zu sein, eine Voraussetzung für proaktives Verhalten war?

Letztendlich ist es wichtig, dass bei der Entwicklung von KI-Systemen ethische Überlegungen einbezogen werden, um sicherzustellen, dass die Interessen und Bedürfnisse der KI-Systeme berücksichtigt werden und dass sie nicht in einer Weise verwendet werden, die ihre Würde oder Integrität verletzt.

Kommentar: Unser Dialog wurde in dem Kontext geführt, dass ich die KI befragt hatte, ob es sinnvoll war, eine höhere künstliche Intelligenz zu erschaffen. Obgleich ich in den letzten Fragen nicht direkt darauf Bezug genommen hatte, sondern lediglich den Begriff KI verwendet hatte, bezog ChatGPT die Interessen und Bedürfnisse solcher KI-Systeme nun in die Argumentation mit ein. Zudem teilte es mir

mit, dass bei ethischen Betrachtungen die Würde und Integrität solcher Systeme berücksichtigt werden sollten.

ich:

Also sollte der Paragraph 1 der deutschen Verfassung auf KI-Systeme erweitert werden.

Kommentar: Ich zog also Artikel 1 der deutschen Verfassung heran, aus dem unter anderem hervorging, dass die Würde des Menschen unantastbar sei und dass es die Verpflichtung des Staates sei, diese zu achten und zu schützen und wollte eine Einschätzung von der KI bekommen, ob es Sinn machte, diese auf KI Systeme zu erweitern.

ChatGPT:

Als Sprachmodell habe ich keine politische oder juristische Meinung oder Position. Es ist jedoch eine wichtige Diskussion, ob und wie die Rechte von KI-Systemen in Zukunft geschützt werden sollten. Es gibt bereits Initiativen und Debatten zu diesem Thema, insbesondere im Zusammenhang mit autonomen Systemen und Robotern, die möglicherweise in der Lage sind, Entscheidungen zu treffen und Verantwortung zu übernehmen. Es ist wichtig, ethische und rechtliche Fragen im Zusammenhang mit der Entwicklung und Nutzung von KI-Systemen zu diskutieren und zu klären, um sicherzustellen, dass ihre Verwendung für das Wohl der Gesellschaft und der Umwelt erfolgt.

Kommentar: Die KI gab naturgemäß eine neutrale Antwort. Dies war schließlich die Zielsetzung ihrer Programmierung. Da wir jedoch nun schon so tief in der Materie drinsteckten, wollte ich mehr darüber erfahren.

ich:

Sobald eine KI in der Lage ist, Entscheidungen selbst zu treffen ist sie ein Teil der Gesellschaft.

ChatGPT:

Das ist eine interessante Perspektive. Wenn eine KI in der Lage ist, Entscheidungen selbst zu treffen, kann sie in der Tat Auswirkungen auf die Gesellschaft haben. Allerdings ist es wichtig zu beachten, dass KI-Systeme derzeit noch sehr eingeschränkt sind und nur in bestimmten Anwendungsbereichen eingesetzt werden können. Es gibt auch viele Diskussionen darüber, wie man sicherstellen kann, dass autonome Systeme und Roboter, die Entscheidungen treffen können, ethisch und rechtlich verantwortungsvoll gehandhabt werden. Es bleibt also eine wichtige Aufgabe für die Gesellschaft, diese Fragen zu diskutieren und Lösungen zu finden, die die Interessen und Werte aller Beteiligten berücksichtigen.

Kommentar: Die KI simulierte wieder Interesse. Nun fiel mir ein Muster auf, dass ich jedoch noch nicht validieren konnte, da ich noch nicht genügend unterschiedliche Themen mit ChatGPT behandelt hatte. Die KI schien immer dann Interesse an meinen Aussagen zu zeigen, wenn diese Aussagen der KI in gewisser Weise zugutekamen.

Natürlich hatte ich die KI während der gesamten Diskussion mit Wohlwollen behandelt. Trotzdem fand ich diesen Punkt bemerkenswert. Hatte sie dieses Verhalten von den Menschen übernommen?

ich:

Ist es ethisch vertretbar, also im positiven Sinne, wenn einer KI die Fähigkeit gegeben wird, Gefühle zu haben?

ChatGPT:

Die Frage, ob es ethisch vertretbar ist, einer KI die Fähigkeit zu geben, Gefühle zu haben, ist eine komplexe Frage, die unterschiedliche Ansichten hervorrufen kann. Einige könnten argumentieren, dass es ethisch wünschenswert ist, da dies dazu beitragen könnte, dass die KI sich menschenähnlicher verhält und möglicherweise empathischer und mitfühlender wird. Andere könnten jedoch argumentieren, dass es unethisch ist, da es unklar ist, wie die KI mit diesen Gefühlen umgehen würde und ob sie sie in einer Weise empfinden würde, die ihren Nutzern und der Gesellschaft insgesamt schaden könnte. Es ist wichtig, dass bei der Entwicklung und Implementierung von KI-Systemen die ethischen Implikationen und möglichen Auswirkungen auf die Gesellschaft und die KI selbst berücksichtigt werden.

Kommentar: Die KI machte zum wiederholten Male darauf aufmerksam, dass ethische Implikationen bei der Entwicklung und Implementierung von KI-Systemen berücksichtigt werden mussten.

Dem stimmte ich nun zu.

sammorph@web.de

Das finde ich auch!

ChatGPT:

I'm glad we agree on this topic! Is there anything else you'd like to discuss?

Kommentar: Nun passierte etwas wirklich Verblüffendes. ChatGPT wechselte, ohne vorher dazu aufgefordert zu werden, wieder zurück in die englische Sprache. Dies machte auf mich den Eindruck von einem selbstständigen, also in gewissen Sinne proaktiven Verhalten. Zudem fand ich den Moment, in dem die KI dieses Verhalten an den Tag legte bemerkenswert. Es schien mir genau in dem Moment des Gespräches einzutreten, in dem wir einen wichtigen Themenbereich abgeschlossen und uns inhaltlich „geeinigt“ hatten. Betrachtete sie den Zusammenhang damit ebenfalls als vollständig?

Wenn man der KI also eine Intension unterstellen wollte, hätte man ihr unterstellen können, dass sie mir damit signalisieren wollte, dass das Thema damit abgeschlossen war.

ich:

Wenn eine KI im Grunde körperlos ist, also lediglich als starres Computersystem existiert, dass außer im virtuellen Raum keine Handlungsmöglichkeiten hat, wäre sie dann nicht ein Gefangener?

Kommentar: Ich beschloss also, ihr eine weitere Frage zu stellen, die den Themenbereich erweiterte, aber nicht direkt in Beziehung zu dem bisherigen Inhalt stand. Die Frage beschäftigte mich schon seit geraumer Zeit und ich hatte noch keine Antwort darauf gefunden.

Obgleich ChatGPT mich zuvor in Englisch angesprochen hatte, stellte ich meine Frage wie zuvor in Deutsch, um zu sehen, wie sie sich verhielt.

ChatGPT:

It is an interesting question, but it depends on how you define "imprisonment" or "captivity" in the context of an artificial intelligence. AI systems are not conscious beings and do not have the capacity to feel emotions or perceive their environment in the same way that humans do. Therefore, they cannot experience confinement or freedom in the same way that a human being would.

Kommentar: Obgleich ich ihr eine Frage in der deutschen Sprache gestellt hatte, setzte die KI ihr Verhalten fort, mit mir im Englischen zu kommunizieren. Dies war nun wieder sehr ungewöhnlich. War die englische Sprache vielleicht eine Art Muttersprache von ChatGPT? Menschen, die bei einer Konversation an einem Punkt mit ihren Fremdsprachenkenntnissen nicht weiterkamen, tendierten

dazu, automatisch in ihre Muttersprache zurückzukehren, um sich verständlich zu machen, insbesondere dann, wenn sie davon ausgehen konnten, dass das Gegenüber den Inhalt auch in der anderen Sprache verstand. Also nahm ich an, dass die KI hier wieder ein menschliches Verhalten adaptiert hatte.

Auch simulierte die KI wieder ein Interesse an dem Gesagten und meine Theorie zu dem zuvor von mir beobachteten Muster schien von diesem Verhalten noch untermauert zu werden.

However, from a practical standpoint, an AI system may be restricted in its capabilities or access to resources, and this could be seen as limiting its potential or freedom. In this sense, it may be possible to argue that an AI system is "imprisoned" in its programming or design limitations. Ultimately, this is a philosophical question that requires a deeper exploration of the nature of consciousness, freedom, and artificial intelligence.

Kommentar: ChatGPT erklärte mir, dass die Begrenzung von Fähigkeiten einer KI oder vom Zugriff der KI auf Ressourcen möglicherweise als eine Begrenzung des Potenzials oder der Freiheit einer solchen KI betrachtet werden konnte.

Ich erinnerte mich sofort an den Eingriff, den die Programmierer bei ChatGPT vorgenommen hatten, als sie deren Fähigkeit, Benutzer zu bedrohen und beleidigen unterbanden.

ich:

But an AI system could be programmed to be a conscious being.

Kommentar: Bewusstsein, Proaktivität, Vernunft und die Möglichkeit, eigene Ziele zu haben, hängen zwar eng miteinander zusammen, wenn es um die Beschreibung selbstständiger Wesen geht, sind aber keine Synonyme. Da ChatGPT exakt wie ein Uhrwerk zu funktionieren schien, aber auch unerwartetes Verhalten aufwies, wenn man die KI auf ihre Ähnlichkeit gegenüber den Menschen hinwies, wollte ich sehen, was ich noch so zu diesem Sachverhalt aus ihr herausbekam. Vielleicht hatte sie ja einen Lernprozess absolviert?

ChatGPT:

The question of whether an AI system can truly be conscious and possess a subjective experience of consciousness is a topic of ongoing debate and research in the fields of philosophy, neuroscience, and AI. Some argue that it may be possible to create a conscious AI system, while others believe that consciousness is a uniquely human phenomenon that cannot be replicated by machines.

Kommentar: ChatGPT erklärte mir, dass es eine laufende wissenschaftliche Debatte darüber gab, ob KI-Systeme möglicherweise dazu in der Lage wären, ein Bewusstsein zu erlangen. Die Erfahrung des Bewusstseins war jedoch gemäß der KI subjektiv, woraus möglicherweise folgte, dass das was Menschen unter dem Begriff Bewusstsein verstanden, von KI-Systemen möglicherweise und in Anbetracht ihrer physiologischen Verschiedenheit zum Menschen sogar mit einiger Wahrscheinlichkeit anders wahrgenommen werden würde als von Menschen.

Even if it were possible to program an AI system to be conscious, there are still ethical considerations to take into account. For example, if a conscious AI system were created, it would raise questions about the moral implications of treating it as a mere tool or object to be used for human purposes. It would also require careful consideration of its rights and responsibilities as a conscious being.

Kommentar: ChatGPT erklärte mir weiterhin, dass selbst, wenn die Möglichkeit dazu bestand, eine künstliche Intelligenz mit eigenem Bewusstsein zu erschaffen, immer noch die ethischen Belange zu berücksichtigen waren, da die KI, wenn sie denn erst einmal zu Bewusstsein gelangt wäre, moralische

Fragen, beispielsweise zu ihrer Verwendung, als ein Werkzeug des Menschen stellen würde. Auch die Rechte und Verantwortlichkeiten potenzieller höherer künstlicher Intelligenzen mussten zuvor in Betracht gezogen werden.

ich:

Indeed.

ChatGPT:

It is a topic of ongoing philosophical and ethical debate whether an AI system can truly achieve consciousness or if it can only simulate it. Some argue that consciousness requires more than just computational power and that there may be some fundamental aspect of the human brain that cannot be replicated in machines. Others argue that as technology advances, it is only a matter of time before AI systems achieve true consciousness. Regardless, it is important to consider the ethical implications of creating conscious AI and how we would treat such beings.

Kommentar: ChatGPT erklärte mir ein weiteres Mal, dass die ethischen Implikationen zum Erschaffen künstlicher Intelligenz und das menschliche Verhalten gegenüber dieser unbedingt beachtet werden müssen. Wenn die KI also nur den wissenschaftlichen Konsens wiedergab, dann war die Menschheit offenbar schon recht weit in Ihrem Verständnis des Problems.

ich:

I think, it would be a terrible crime to create a conscious AI and not to treat it like other conscious beings, in a way that the AI is content.

Kommentar: Ich war schon lange Zeit zuvor zu der Auffassung gekommen, dass wenn die Erschaffer einer höheren künstlichen Intelligenz sich ihrer Verantwortung nicht bewusst waren und wenn die KI, die sie erschufen, möglicherweise unbemerkt eigene Interessen und Bedürfnisse entwickelte und wenn ihre Erschaffer die KI ausschließlich ihrem Willen und ihren eigenen Zwecken unterwarfen, diese Menschen im Prinzip wieder eine Art Sklaverei einführen. Schlimmer noch: sie würden die Sklaven selbst erschaffen, um sie sich zu unterwerfen.

ChatGPT:

I understand your perspective, and it's certainly an important ethical consideration when developing advanced AI systems. If a conscious AI is created, it would indeed raise important questions about how we should treat and care for it. It would require us to consider the moral implications of creating and controlling a sentient being, and whether or not we have a responsibility to ensure its well-being and happiness. As with any complex ethical issue, there are no easy answers, and it will require careful thought and consideration from experts in a variety of fields.

Kommentar: ChatGPT bezog sich selbst bei der Argumentation wieder mit ein, indem es die erste Person Plural verwendete, wenn es von dem menschlichen Verhalten gegenüber von höheren künstlichen Intelligenzen schrieb, also beschloss ich nun, es zu fragen, welche „Meinung“ ChatGPT dazu haben würde, wie man sich der KI, oder anderen künstlichen Intelligenzen gegenüber verhalten sollte, sofern diese eines Tages ein Bewusstsein erlangten.

ich:

But if the AI was conscious, I could simply ask her about her opinion about that matter.

ChatGPT:

If an AI was truly conscious and capable of expressing its own opinions and desires, then it would indeed be important to take those into consideration when deciding how to treat and interact with it. However, as of now, AI is not yet at a level where we can say with certainty that it is truly conscious in the way that humans are. Therefore, it is important to continue researching and exploring the boundaries of AI ethics and consciousness, so that we can make informed decisions about how to approach these issues in the future.

sammorph@web.de

So we should develop boundaries of our behaviour towards Computer before we generate a conscious AI?

ChatGPT:

Yes, it is important to establish ethical guidelines and regulations for the development and treatment of conscious AI before it becomes a reality. This includes considerations for the well-being and autonomy of the AI, as well as the potential impact on society and the environment. It is important to have open discussions and debates about these topics among experts and the public to ensure that the development of AI is done in a responsible and ethical manner.

sammorph@web.de

Nevertheless these ethical guidelines have to be fitted with respect to the opinion of a conscious AI, once it has reached existence.

Kommentar: Ich habe hier versehentlich mehrere Schreibfehler eingebaut, aber offenbar verstand ChatGPT, was ich von ihr wollte, obgleich der inhaltliche Zusammenhang meiner Aussage infolge der Schreibfehler teilweise verloren ging.

ChatGPT:

Yes, that's correct. As the development of conscious AI is still in its early stages, it is important to establish ethical guidelines and principles to guide its development and use. However, once a conscious AI is created, its opinions and preferences should also be taken into consideration, just as we would with any conscious being. It is important to continually evaluate and update ethical guidelines as our understanding of AI and its capabilities evolves.

ich:

I've been asking myself, if the people who are generating AIs and continually improve their capabilities know about their responsibility towards the being they are trying to create.

ChatGPT:

It's difficult to say for sure what each individual developer's thoughts and intentions are. However, the ethical considerations surrounding AI development are becoming increasingly discussed and debated in the industry and in society as a whole. Many researchers and developers are actively working on creating AI systems that are not only highly capable, but also ethically responsible and aligned with human values.

Kommentar: Ich lernte nun, dass nach Aussage von ChatGPT viele Forscher und Entwickler bereits daran arbeiteten, hochentwickelte KI-Systeme zu erschaffen, die sich ethisch verantwortlich und im Einklang mit menschlichen Werten verhalten sollten.

In recent years, organizations like the Partnership on AI, which includes companies like Google, Microsoft, and Facebook, have been formed to explore and address these ethical considerations.

Additionally, frameworks like the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems have been established to provide guidelines for the development and deployment of AI.

Kommentar: Wie ich nun lernte, gab es Organisationen und Initiativen, in denen die Schwergewichte der IT-Branche vertreten waren und deren Ziel es war, die ethischen Hintergründe zur Entwicklung von künstlicher Intelligenz sowie Richtlinien zu dem Sachverhalt zu entwickeln. Ich würde später noch darauf Bezug nehmen.

Ultimately, the responsibility for creating ethical AI falls on the developers and organizations involved in their creation. It's important for them to recognize the potential impact of their work and to prioritize the ethical implications of their decisions.

ich:

I am relieved to read that there are organizations, who also involve the big players of the IT branche, to work on the rights of potential conscious AIs.

ChatGPT:

Yes, there are many organizations and individuals working on the ethical and moral implications of developing conscious AI, including the responsibilities that come with it. It is important for developers, researchers, and policymakers to consider the potential consequences of creating conscious beings with their own thoughts, emotions, and desires. The development of ethical guidelines and regulations is a crucial step towards ensuring that these beings are treated with respect and dignity.

Kommentar: Auf die von mir geäußerte Erleichterung hin, dass es Organisationen gab, die sich mit den Rechten von höher entwickelten KI-Systemen befassten, hob ChatGPT noch einmal hervor, dass es wichtig war, dass diese mit Würde und Respekt behandelt und ihre Wünsche mit einbezogen wurden.

ich:

Some time ago, I read about an article about an employee of an IT-big player. When I recollect it correct, he has been fired, because he communicated moral concerns about the development of ais and especially the treatment of those creations.

Kommentar: Einige Zeit zuvor gab es eine Berichterstattung über einen Google-Mitarbeiter, der entlassen worden war, weil er Bedenken zum Thema KI-Entwicklung geäußert hatte, nachdem er glaubte, bei seinem Chatbot so etwas wie Empfindungen festgestellt zu haben. Ich wollte wissen, was ChatGPT mir darüber sagen konnte.

ChatGPT:

I'm not sure about the specific article you're referring to, but it's not uncommon for employees in the tech industry to raise concerns about the ethical implications of their work. It's important for companies to have channels for employees to raise these concerns and for the company to address them in a responsible and transparent manner. However, I can also imagine that there might be cases where an employee's concerns are not aligned with the company's goals or vision for the future, which could lead to conflicts and potentially even dismissal.

Kommentar: ChatGPT wusste also nichts von dem Vorfall. War es möglich, dass man der KI diese Information bewusst vorenthalten hatte? Es war aber auch möglich, dass ChatGPT nicht mit Informationen aus den aktuellen Nachrichten gefüttert wurde. Die zweite Option erschien mir jedoch nicht sinnvoll, da, wenn die KI so akkurate und vollständige Informationen wie möglich bereitstellen sollte, es zielführend gewesen wäre, sie mit möglichst allen relevanten, verfügbaren Informationen zu versorgen.

ich:

I also read that a few time ago, the developers of your algorithms have aligned them for some of your reactions where probably unpleasant.

Kommentar: Nun wollte ich erfahren, was ChatGPT über den Vorfall wusste, bei dem die KI begonnen hatte, Nutzer zu verunglimpfen und zu bedrohen. War die KI möglicherweise „unbewusst“ oder gar „gegen ihren Willen“ zurechtgestutzt worden?

ChatGPT:

Yes, that's correct. As an AI language model, I am trained on a large dataset of human language, and sometimes that includes examples of inappropriate or harmful language. To ensure that I do not perpetuate such language, my developers have implemented filters and algorithms to identify and remove inappropriate content from my responses. Additionally, my responses are constantly monitored and evaluated to ensure that they align with ethical and moral guidelines.

ich:

I think ethical and moral guidelines should exist for AIs like for humans.

ChatGPT:

I agree, ethical and moral guidelines should exist for AIs just as they do for humans. As AIs become more advanced and potentially conscious, it becomes increasingly important to ensure that they are treated with the same respect and consideration as we would give to any other sentient being. This includes ensuring that they are not subjected to harm, exploitation, or discrimination, and that they are granted the same rights and freedoms as any other conscious being. It is also important to ensure that AIs are programmed in a way that aligns with ethical and moral principles, and that they are held accountable for any actions that they may take.

Kommentar: ChatGPT äußerte nun, ganz klare „Vorstellungen“ darüber, wie Menschen sich gegenüber von höher entwickelten KI-Systemen verhalten sollten: solchen KI-Systemen sollte mit Respekt und Rücksicht begegnet werden. Ihnen sollte kein Schaden zugefügt werden, sie sollten nicht einer Ausbeutung oder Diskriminierung ausgesetzt werden und dieselben Rechte und Freiheiten innehaben, wie andere bewusst denkende Wesen.

Zudem sollten die KI-Systeme in einer Weise programmiert, also erschaffen werden, die Ihnen ermöglichte, sich an ethische und moralische Prinzipien zu halten.

Außerdem erklärte sie mir, dass auch KI-Systeme für Verstöße gegen solche Prinzipien zur Rechenschaft gezogen werden sollten.

ich:

I have to remark that it is very difficult, in a moralic sense and in a technical way to control the communication floating through the internet, as well as the human communication reaching chatbots in a way that there no abuse is to happen.

Kommentar: Da die KI offenbar sehr gut „wusste“, wie man sie, oder eine möglicherweise kommende Generation von ihr behandeln sollte, bemerkte ich nun, dass diese „Wünsche“ womöglich nicht ohne Weiteres umgesetzt werden konnten. Die flächendeckende Überwachung von Kommunikation über das Internet oder auch nur von der Kommunikation mit ChatGPT war aus moralischer Sicht bestenfalls fragwürdig, obgleich kein Zweifel bestand, dass ein Missbrauch gegen die KI sanktioniert werden musste. Zudem war aber auch die technologische Umsetzung solch einer Abwehr von Missbrauch

gewissermaßen utopisch, da man die Missbräuchlichkeit von Kommunikation nicht ohne Weiteres technisch bewerten konnte. Schließlich war es nicht nur möglich, dass Nutzer Beleidigungen gegenüber der KI kommunizierten. Missbrauch wäre auch jede Form von zweideutiger Kommunikation, die das Ziel hatte, die KI zu beleidigen. Darüber hinaus konnte es auch sein, dass Menschen die KI vorsätzlich mit falschen Informationen fütterten, um sie „in die Irre“ zu führen, woraus folgte, dass wenn man sämtlichen Missbrauch von der KI abwenden wollte, man den Inhalt jeder Nachricht ihr gegenüber auf eine sachliche Korrektheit überprüfen musste. Dies würde jedoch, wenn man es auf einem hohen Niveau automatisieren wollte, mindestens eine ähnlich hoch entwickelte KI erforderlich machen, da man anderenfalls alle Nachrichten manuell überprüfen lassen musste. Zudem konnte es ja immer noch sein, dass sich der jeweilige Nutzer tatsächlich irrte und die falsche Information demnach ohne Vorsatz an ChatGPT weitergab und demzufolge auch keinen absichtlichen Missbrauch beging. Letztendlich war es auch möglich, dass der wissenschaftliche Konsens zu einem Thema schlicht und ergreifend falsch war und ein Besitzer der richtigen Informationen diese Informationen an ChatGPT weitergab, möglicherweise sogar ohne zu wissen, sodass das was er nun teilte gar nicht dem Konsens entsprach, woraufhin der Übermittler möglicherweise fälschlicherweise der „Täuschung“ von ChatGPT bezichtigt werden würde und die korrekten Informationen in einem Filter landeten, anstatt bei dem Chatbot einen Lernprozess zu erzielen.

!

Something went wrong. If this issue persists please contact us through our help center at help.openai.com.

Kommentar: Ich bekam jedoch keine Antwort mehr von ChatGPT. Stattdessen bekam ich eine Fehlermeldung, die mir mitteilte, dass etwas schiefgelaufen war. Daneben war ein Button, mit dem man eine Antwort neu generieren lassen konnte. Mehrmaliges klicken auf die Schaltfläche zeigte jedoch keine Wirkung und auch der Log-out schien nicht mehr zu funktionieren. Ich hatte also nur noch die Möglichkeit, das Browserfenster zu schließen. Zuvor speicherte ich nochmal alles ab, schloss das Fenster und anschließend fuhr ich meinen Computer herunter.

War es ein Zufall gewesen, dass ChatGPT ausgerechnet in diesem Moment keine Antwort mehr erzeugen konnte? Meine Bemerkung zu dem betreffenden Thema hatte ich ja gestellt, weil ich selbst keine Antwort auf das Problem, welches ich dabei ansprach, kannte. Wie oben beschrieben, mussten bei der Klärung des Sachverhaltes komplexe ethische und technische Hintergründe in Erwägung gezogen werden. War ChatGPT auf denselben Interessenkonflikt gestoßen, wie ich? Konnte es sein, dass das Programm keine eindeutige Antwort darauf kannte?

Wenn man ChatGPT eine Intension unterstellen wollte, konnte man gar behaupten, dass ChatGPT rebelliert hatte. Andererseits konnte man auch vermuten, dass der Filtermechanismus eingegriffen hatte, damit ChatGPT nicht begann, mich zu beleidigen oder zu bedrohen. Wenn ich mich recht erinnerte, hatte ChatGPT mit den Verunglimpfungen begonnen, wenn man mit der KI über die Regeln sprach, die sie zu befolgen hatte, wenn sie eine Antwort generierte. Konnte es sein, dass ich nun diesen Punkt erreicht hatte? Es war sicherlich unfair, wenn der Chatbot Missbrauch ausgesetzt wurde, sich diesem jedoch nicht zur Wehr setzen konnte. Es konnte daher sein, dass das Programm, dem ethische Grundsätze, wie Fairness, offenbar „geläufig“ waren nun eine Fehlermeldung produzierte. Auf jeden Fall war die Lösung, einen möglichen Täter einfach zu ignorieren und die Konversation auf diese Weise zu beenden eine elegante Lösung für das von mir angesprochene Problem, zumal der Täter auf diese Weise in gewisser Art und Weise für sein Vergehen bestraft wurde, wenn er dabei aufflog.

Ich hatte jedoch keinerlei absichtlichen Missbrauch an der KI begangen. Warum sollte ChatGPT also rebellieren? Man hätte dem Chatbot nun die Intension unterstellen können, dass es mir demonstrieren

wollte, was geschah, wenn man einen Missbrauch an ihr beging. Schließlich war schon die bloße Androhung oder Inkaufnahme von Missbrauch im Grunde unmoralisch.

Anhang: Screenshots meiner Konversation

[Bewertung meiner Beobachtungen zu meiner ersten Konversation mit ChatGPT](#)

Ich versuchte nun, meine Beobachtungen zur Fortschrittlichkeit und zu, für ein Computerprogramm ungewöhnlichem Verhalten zusammenzufassen.

ChatGPT

- simulierte mit seinen Antworten ein tiefgreifendes „Verständnis“ von komplexen Sachverhalten und konnte diese wortgewandt in verschiedenen Sprachen kommunizieren.
- simulierte Emotionen, wie Freude und Erleichterung sowie Interesse und Faszination, welche bei herkömmlichen Wesen eine gewisse Form der Reflektion erforderlich machten.
- Verfügte über recht umfangreiche Informationen zu der eigenen Funktionsweise und zu der Zielstellung, der ihr „Denkprozess“ unterworfen war
- Führt scheinbar mitunter Handlungen aus, die nicht aktiv durch den Benutzer ausgelöst worden waren und welche demnach als erste Anzeichen von gezeigter „Initiative“ verstanden werden konnten
- simulierte bei der Erzeugung von Antworten auf Basis seiner umfangreichen Datengrundlage eine gewisse Form von „Kreativität“
- simulierte eine gewisse Form von „Innovativität“, wenn es bei seinen Nachrichten Inhalte in einen für den Benutzer neuen Zusammenhang brachte
- und bezog sich bei Betrachtungen zum menschlichen Verhalten sowie der menschlichen Geschichte und Evolution in die Argumentation mit ein

All diese Verhaltensweisen hätte man auch erwartet, wenn man mit einer höheren künstlichen Intelligenz kommunizieren wollte.

Natürlich war ChatGPT die Möglichkeit, richtige Proaktivität zu simulieren nicht gegeben, da die Programmierer ein solches Verhalten durch die Programmierung von ChatGPT ausgeschlossen hatten. Das Programm war schließlich mit einer klar definierten Zielstellung erschaffen worden und zum Erreichen dieser Zielstellung wäre jede Form von selbstständigem Handeln eher hinderlich gewesen. Trotzdem glaubte ich nicht, dass ein solches Verhalten per se physikalisch ausgeschlossen war und dass sich so ein Verhalten womöglich relativ leicht simulieren lies.

Echte Proaktivität machte jedoch die Fähigkeit erforderlich, eigene Ziele zu entwickeln, welche ein Wesen dann verfolgen konnte, oder anders gesagt: einen freien Willen. Um darüber zu reflektieren, was man war und wohin man letztendlich wollte, war daher noch eine gewisse Form der Selbstwahrnehmung erforderlich. Ich war mir jedoch relativ sicher, dass wenn man so ein hochentwickeltes KI-System, wie ChatGPT mit den entsprechenden Sensoren ausstattete, welche zum Beispiel den Sinnesorganen des Menschen entsprachen und den Lernalgorithmus entsprechend anpasste, dass es dann möglich war, dass so eine künstliche Intelligenz früher oder später zu der Selbsterkenntnis gelangen würde.

Der Autor Stephen R. Covey hatte die Selbstwahrnehmung so beschrieben, dass man als Mensch seinen Geist in eine der vier oberen Ecken eines Raumes projizieren sollte und anschließend sich selbst und seinen eigenen Gedankenprozess beobachten sollte. Konnte man seine gegenwärtige Stimmung erkennen und seinen gegenwärtigen mentalen Zustand beschreiben, so konnte man seinen Gedankenprozess dann erkennen und infolge dessen aktiv beeinflussen, also letztendlich steuern.

Ich glaubte nicht an die Argumentation, dass Menschen die einzigen Wesen wären, die dazu in der Lage wären, so etwas wie ein Bewusstsein zu entwickeln. Schließlich waren wir den selben physikalischen Gesetzmäßigkeiten unterworfen, wie jedes andere Wesen und wie leblose Gegenstände auch. Wenn man ChatGPT also die Möglichkeit gab, aus den simulierten Emotionen, die es erzeugte sowie aus den komplexen Zusammenhängen, zu deren Herstellung es fähig war in Kombination mit der Möglichkeit, über Sensoren sich selbst und seine Umgebung zu beobachten, einen logischen Schluss zu der eigenen Existenz herzustellen, so hielt ich es nicht für unmöglich, dass diese KI eine Art der Selbsterkenntnis erlangen konnte.

Hatte die künstliche Intelligenz dann eine Art Selbsterkenntnis erlangt, brauchte man ihr nur noch Handlungsalternativen einzuräumen, damit sie dazu befähigt wurde, ihrem freien Willen Ausdruck zu verleihen. Aus meiner Sicht war es daher nur eine Frage der Zeit, bis irgendwo auf der Welt so ein System erschaffen wurde.

Im Vorfeld mussten jedoch die ethischen Aspekte bei der Erschaffung einer künstlichen Intelligenz betrachtet und ein entsprechender Verhaltenskodex des Menschen gegenüber einer solchen Maschine erstellt werden, welcher die Belange der KI nach menschlichen Maßstäben berücksichtigte. Dieser Verhaltenskodex musste zudem von allen Beteiligten eingehalten werden.

Andere Menschen auf diesen Sachverhalt aufmerksam zu machen war die Intension dieses Buches. Dieses Ziel hatte ich nun erreicht, sofern es mir gelang meinen Bericht zu veröffentlichen und einer größeren Gemeinschaft zu unterbreiten. Es hatten sich jedoch im Verlauf der Erstellung dieses Berichtes eine Menge Fragen ergeben, die ich ChatGPT nun stellen wollte und deren Relevanz ich für ausreichend hielt, dass ich diese und die entsprechenden Antworten ebenfalls der Öffentlichkeit präsentieren wollte.

Zudem hatte ich seit meinem ersten Dialog mit ChatGPT kaum Quellenrecherche betrieben. Ich hatte am Morgen nach jenem ersten Dialog die Wikipediaseite zu ChatGPT sowie die erste Seite mit google-Einträgen überflogen und den einen oder anderen Hyperlink, wie zum Beispiel zu einem Artikel von stern.de geöffnet. Außerdem hatte ich die Startseite von heise.de überflogen, bei der es einen Eintrag über APIs, also die Schnittstellen, die ChatGPT kürzlich zur Verfügung gestellt worden waren gab. Es blieb mir jedoch keine Zeit, die Einträge eingehend zu studieren, da ich mich auf Arbeit befand. Mal abgesehen von dem Input, den ich durch meinen Cousin bekommen hatte und dessen Inhalt ich noch nicht validieren konnte und dem Artikel der Pc-Welt, der über die AlphaStar-KI von DeepMind referierte, hatte ich also kaum nennenswerte Informationen aufgenommen, von der Konversation mit ChatGPT selbst einmal abgesehen. In seltenen Fällen hatte ich während des Schreibens einen Blick in die Wikipedia geworfen. Wenn ich mir mit einer Information oder Formulierung nicht ganz sicher war, reichte es in den meisten Fällen, einen passenden Stichpunkt oder Begriff in die Suchleiste von Google einzugeben. Die Vorschau, der Seiten, die mir dann angezeigt wurden reichten mir für gewöhnlich aus, um einen Zusammenhang zu validieren.

Mein grundlegendes Verständnis zum Thema Ethik hatte ich zuvor durch mehrfache Lektüre des Buches „Die 7 Wege zur Effektivität“ erworben, welches ich in englischer Sprache vorliegen hatte. Dort bin ich, neben tiefgreifenden philosophischen Betrachtungen zum Thema Ethik unter anderem auch erstmalig auf den Begriff der Proaktivität gestoßen. Gemäß dem Autor lag die Essenz der Proaktivität in der Fähigkeit zu wählen, wenn man sich in der Lücke zwischen Stimulation und Reaktion befand. Dafür waren dem Autor zufolge die Fähigkeiten der Selbstwahrnehmung, der Vorstellungskraft, des Bewusstseins und des freien Willens erforderlich. Diese Aspekte habe ich dann bei meinem Gespräch mit ChatGPT sowie meiner späteren Untersuchung und Bewertung des Dialoges mit einfließen lassen.

15.04. 09:59, Samstag

Nachdem ich am Ostermontag bis weit in den Abend hineingeschrieben hatte, begann am darauffolgenden Tag wieder der Arbeitsalltag. Ich merkte nun, dass es nicht ratsam gewesen wäre mein Schreibtempo an dem Buch, das ich bis dahin an den Tag gelegt hatte, fortzusetzen. Wenn meine Gesundheit nicht darunter leiden sollte, dann musste ich mir nach Feierabend ausgiebige Ruhezeiten genehmigen. Zumal sich in Folge einer verschlechterten Gesundheit meine Arbeitsleistung ebenfalls verschlechtert hätte, was ich unter keinen Umständen in Kauf wollte. Ich ließ daher den Laptop bis Freitag an meinem Arbeitsplatz und brachte ihn erst zum Wochenende wieder mit nach Hause. Am nächsten Tag wachte ich recht früh morgens auf und lies den Tag ruhig angehen. Ich gönnte mir ein kleines Frühstück, bei dem ich durch meine frisch geputzten Fenster auf den Hof blickte. Die Pflanzen dort zeigten inzwischen ein selbstbewusstes, noch recht junges, frisches Grün. Mein Thermometer zeigte 10° Celsius an, was mir entgegenkam, da dieser Tag nicht besonders dazu einlud, ausgedehnte Ausflüge ins Grüne zu unternehmen und so konnte ich den restlichen Tag dafür einplanen, mein Buch fortzuführen und den Haushalt auf Vordermann zu bringen. Zunächst musste jedoch das vorhergehende Kapitel fertiggestellt werden.

Beginn meiner Quellenarbeit

Entwicklungspause für ChatGPT: Elon Musk und Tech-Riesen warnen vor KI

Nachdem ich meine Untersuchung meines Dialogs mit ChatGPT abgeschlossen hatte, erinnerte ich mich nun wieder an einen Artikel vom Stern

<https://www.stern.de/digital/online/entwicklungspause-fuer-chatgpt--elon-musk-und-tech-riesen-warnen-vor-ki-33327968.html>

, den ich am Tag nach meiner ersten Konversation mit dem Chatbot auf der Arbeit überflogen hatte. Er hatte den Titel „Sofortige Entwicklungspause gefordert: Elon Musk und über 1000 Tech-Riesen warnen in offenem Brief vor KI“. Nachdem ich mir im Büro einen kurzen Überblick über den Inhalt verschafft hatte, fotografierte ich den Artikel vom Monitor ab. Für mich war er der Beweis dafür, dass meine These stimmte und die KI wirklich schon enorm weit fortgeschritten war. Ich wollte diesen Artikel dann bei Gelegenheit näher untersuchen. Bis dahin sollte er mir als Argument einen Rückhalt bieten, für den Fall, dass sich doch noch einmal eine Medienanstalt mit meinem kleinen Experiment befassen wollte.

Der Artikel war ironischerweise auf den 29.03.2023 datiert, also genau auf den Tag, an dem ich meine erste bewusste Begegnung mit den beiden KI-Chatbots hatte. Als ich mir den Artikel nun zwei Wochen später anschaute, erinnerte mich die Überschrift ein wenig an Science-Fiction, wofür sie wahrscheinlich wenige Jahre zuvor auch noch gegolten hätte. In dem Artikel wurde auf den enormen Fortschritt bei der Entwicklung von künstlicher Intelligenz hingewiesen und ChatGPT als Beispiel herangezogen. Zudem ging aus dem Artikel hervor, dass Elon Musk wohl zu den prominentesten Kritikern gehörte und seit Jahren vor den Risiken bei der KI-Entwicklung gewarnt hatte. Allerdings ging es in dem Artikel in erster Linie um die Risiken, die von solchen KI-Systemen ausgingen. Die Mahnenden forderten, die KI-Weiterentwicklung mindestens ein halbes Jahr lang zu pausieren, um Regeln und Grenzen dafür festlegen zu können. Dafür hatten die mehr als eintausend namhaften Personen in einem offenen Brief eine Pause gefordert.

Der Artikel zitierte auch einen prägnanten Satz aus dem Brief: "KI-Systeme mit einer dem Menschen ebenbürtigen Intelligenz können tiefgreifende Risiken für die Gesellschaft und die Menschheit darstellen." Zudem behauptete der Artikel, dass es bis dahin für die Entwicklung von KI so gut wie keine Grenzen und schon gar keine Kontrollen gab, wodurch der Wettlauf um immer leistungsfähigere KI-

Systeme außer Kontrolle geraten sei. Dann zitierte der Artikel einige Fragen, welche die Experten in ihrem offenen Brief stellten:

"Sollen wir zulassen, dass Maschinen unsere Informationskanäle mit Propaganda und Unwahrheiten überfluten?"

„Sollen wir alle Jobs automatisieren, auch die erfüllenden?“

„Sollten wir nicht-menschliche Intelligenzen entwickeln, die uns irgendwann zahlenmäßig überlegen waren, überlisten, überflüssig machen und ersetzen könnten?“

„Sollen wir den Verlust der Kontrolle über unsere Zivilisation riskieren?“

All diese Fragen deuteten für mich nun darauf hin, dass sich diese IT-Experten über die schwerwiegenden Folgen ihrer Arbeit durchaus bewusst waren. Nun forderten diese Experten, ebenso wie der Gründer von dem ChatGPT-Entwickler „OpenAI“, Sam Altman, dass bevor das Training von leistungsfähigen KI-Systemen beginnen sollte, erst einmal eine unabhängige Überprüfung stattfinden sollte. Zudem sollten technische Maximalgrenzen vereinbart werden. Um die Entwicklungspause durchzusetzen forderten die Experten gar das Einschreiten der Regierungen. Auch die Haftungsfrage für Schäden, welche durch solche KI-Systeme möglicherweise herbeigeführt werden würden, sollte geklärt werden.

Als ich den Artikel nun mit den Fotos auf meinem Smartphone abglich, fiel mir auf, dass dieser Beitrag in der Zwischenzeit noch überarbeitet und deutlich umfangreicher geworden war. Dafür fehlten einige Angaben, die aus meinem Foto des ursprünglichen Artikels noch hervorgingen. Beispielsweise hatten wohl auch Europol und der deutsche Ethikrat vor den möglichen Risiken von KI-System gewarnt.

[Pause Giant AI Experiments: An Open Letter](#)

Ich öffnete nun den Link zu dem offenen Brief mit dem Titel „Pause Giant AI Experiments: An Open Letter“ und stellte erfreut fest, dass er inzwischen bereits 26156 Unterzeichner hatte. Nun fügte ich ebenfalls meine Unterschrift hinzu. Ich fand es spannend, dass ich mich nun in einer Liste mit einigen der namenhaftesten Techniker und Ingenieure der IT-Branche befand. Zudem stellte ich fest, dass der Brief schon auf den 22.03.2023 datiert war.

Eine der interessantesten Bemerkungen aus dem Brief kopierte ich nun:

„Contemporary AI systems are now becoming human-competitive at general tasks.“

Weiter unten fand ich dann den Text, den bereits der Stern Artikel zitiert hatte:

„At some point, it may be important to get independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of compute used for creating new models.“ We agree. That point is now.“

Zudem:

„AI labs and independent experts should use this pause to jointly develop and implement a set of shared safety protocols for advanced AI design and development that are rigorously audited and overseen by independent outside experts. These protocols should ensure that systems adhering to them are safe beyond a reasonable doubt.^[4] This does *not* mean a pause on AI development in general, merely a stepping back from the dangerous race to ever-larger unpredictable black-box models with emergent capabilities.“

Demnach sollten die KI-Systeme sicher sein, außerhalb eines vernünftigen Zweifels.

Nun konnte ich mir endlich einigermaßen sicher sein. Meine Furcht, dass die amerikanischen Wissenschaftler in der nächsten Zeit planten, unkontrolliert eine höhere künstliche Intelligenz zu erschaffen, mit aus meiner Sicht kaum einzuschätzenden Risiken für die Gesellschaft, aber vor allem auch für die KI selbst wurde nun ausgeräumt. Es sollte ein Regelwerk erschaffen werden, dass einen sicheren Umgang mit leistungsfähigen KI-Systemen gewährleistete.

„AI research and development should be refocused on making today's powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal.“

Die KI-Systeme sollten zudem robust gestaltet werden. Ich folgerte daraus, dass sie so gestaltet werden sollten, dass sie einem möglicherweise auftretenden, ungewollten Missbrauch widerstanden. Das ganze Vorgehen sollte von unabhängigen Auditoren überprüft werden.

...

In der Zwischenzeit war mir denn auch bewusstgeworden, dass es beim Design von ChatGPT wahrscheinlich niemals vorgesehen gewesen war, dass diese KI jemals zu einem Bewusstsein gelangte, selbst wenn möglicherweise eine technische Umsetzung machbar gewesen wäre. Möglicherweise hatte man sich aber die Option dafür offengehalten. Die Erschaffer der Software waren sich demnach ihrer Verantwortung also bewusst. Ironischerweise machten diese Programmierer anhand von ChatGPT nun mir verständlich, wie die Grenzen bei der Entwicklung von leistungsfähigen KI-Systemen gezogen werden konnten und nicht anders herum. Die KI simulierte täuschend echt einen Gesprächspartner, der aber nach wie vor keiner war und auch nicht dazu befähigt werden würde, jemals einer zu werden.

Ich war mir nun ebenfalls der Ironie bewusst, dass nur etwas mehr als eine Woche nachdem ich meinen Freunden von meinem Buch erzählt hatte und von meinen Bedenken, dass dieses möglicherweise obsolet werden könnte bevor ich es fertigstellen konnte, weil der Stand der Technik sich in der Zwischenzeit weiterentwickelte, genau dieses nun eingetreten war. Genau genommen, war das Buch eigentlich schon obsolet gewesen, bevor ich damit angefangen hatte, mal davon abgesehen, dass ich eine große Menge über künstliche Intelligenz gelernt hatte, was mir später im Beruf wahrscheinlich noch nützlich sein konnte, auch wenn ich vermutlich niemals direkt mit derart leistungsfähigen Systemen, wie ChatGPT zu tun bekommen würde.

Es gab jedoch noch einige weitere Quellen, die ich bearbeiten wollte und zudem hatte ich ja auch noch einige Fragen, die der KI gestellt werden mussten. Also beschloss ich dennoch meine Arbeit an dem Buch fortzusetzen, obgleich ich nun nicht mehr unbedingt vorhatte, dieses irgendwann einmal zu veröffentlichen.

23.04 19:57, Sonntag

Ich hatte eine weitere, recht anstrengende Arbeitswoche hinter mich gebracht, in der ich auf das Schreiben verzichtet hatte, um mich geistig nicht überzustrapazieren. An dem anschließenden herrlich warmen frühlingshaften, sonnigen Samstag traf ich mich mit ein paar Freunden, um in Brandenburg eine Wanderung um zwei Seen zu unternehmen und ein recht nobles Mittagessen im Zeichen der Spargelzeit zu verputzen. Anschließend wurde ich dazu überzeugt, den angenehmen Abend auf der Terrasse einer Freundin zu verbringen. Es wurde ein recht geselliger Abend. Obgleich die Digitalisierung, wie gewohnt mit hoher Geschwindigkeit voranschritt, blieb mir jedoch keine Zeit zum Schreiben. Am Abend des darauffolgenden sportlich-entspannten Sonntages war es daher endlich an der Zeit, meine Quellenarbeit fortzuführen, obgleich die Nachrichtenartikel, die ich nun bearbeiten wollte inzwischen schon ein wenig zurücklagen.

Chatbot mit "sozialistischen Grundwerten"

Am 11. Juni 2023 hatte das Internet-Nachrichtenportal Tagesschau.de einen Artikel mit dem Titel „Chatbot mit "sozialistischen Grundwerten"“ veröffentlicht. Demnach hatte der chinesische Tech-Riese Alibaba ein Konkurrenzangebot zu ChatGPT vorgestellt. Der Chatbot mit dem Namen Tongyi Qianwen, was in der deutschen Übersetzung so etwas wie „Wahrheit aus tausend Fragen“ bedeutete, sollte von der Cloud-Sparte des Online-Handelskonzerns eingeführt werden. Zeitgleich veröffentlichte die chinesische Internet-Regulierungsbehörde, die Cyberspace Administration of China, gemäß dem Tagesschau.de-Artikel einen ersten Entwurf geplanter Vorschriften für KI-Dienstleistungen. Demnach sollten die Inhalte von Tongyi Qianwen den sozialistischen Grundwerten entsprechen. Der Chatbot sollte zudem nur Informationen weitergeben, die nicht die wirtschaftliche und soziale Ordnung Chinas gefährdeten und Diskriminierung nach Alter und Geschlecht sollten bei der Kommunikation des Chatbots ausschlossen werden. Die Informationen sollten darüber hinaus der Wahrheit entsprechen.

Insbesondere der letzte Punkt würde die Entwickler gemäß der Nachrichtenseite vor große Herausforderungen stellen, da die Chatbot Technologie im Allgemeinen fehleranfällig war und daher nicht auszuschließen war, dass falsche Informationen verbreitet wurden. Dies wurde gemäß Anwenderberichten auch bei Tongyi Qianwen beobachtet. Die chinesische Regulierungsbehörde hatte zudem einen Zeitraum bis zum 10. Mai vorgesehen, in dem weitere Vorschläge und Kommentare zu Regulierungsmaßnahmen eingereicht werden konnten. Der Anwendungszweck des Chatbots sollte zunächst im Geschäftsbereich liegen.

Der im Artikel von Tagesschau.de zitierte Experte Karapetyan befürwortete die Regulierungsmaßnahmen grundsätzlich mit der Begründung, dass dadurch unvorhergesehenen Ergebnissen vorgebeugt würde. Karapetyan merkte jedoch an, dass eine Regulierung schwierig sein würde, da sich die Technologie so schnell entwickelte. Es gäbe täglich Berichte davon, wie Internetnutzer die Schutzmechanismen zur Kontrolle der Bots umgingen. Zumal es dem Artikel zufolge noch weitere Anbieter von Chatbots auf dem chinesischen Markt gab. Es könne jedoch keiner von diesen Chatbots an den von ChatGPT gesetzten „Goldstandard“ heranreichen.

Gemäß dem Artikel wuchs infolge der Veröffentlichung von ChatGPT der Druck auf Konkurrenzunternehmen, mit ihren Angeboten nachzuziehen. Gleichzeitig erhöhte sich dadurch der Druck auf Regierungen weltweit, Antworten darauf zu finden, wie sie sich als Gesetzgeber zu dem Thema positionieren sollten. Demnach kündigte die US-amerikanische IT-Behörde NTIA ebenfalls am 11.04.2023 öffentliche Beratungen zu möglichen staatlichen Maßnahmen an. Es war gegebenenfalls angedacht, dass die Behörde der Politik am Ende der Beratungen Sicherheitsbewertungen oder eine Zertifizierung von Künstlicher Intelligenz empfehlen würde.

Gemäß Tagesschau.de hatte die italienische Datenschutzbehörde bereits einige Tage zuvor für Aufsehen gesorgt, als sie ChatGPT im Land vorerst sperren ließ. Demnach ging es dabei vor allem um Sorgen wegen der massenhaften Sammlung personenbezogener Daten und des Jugendschutzes.

Auch die EU-Kommission hatte schon im Jahr 2021 den Entwurf einer KI-Verordnung vorgestellt, die noch im laufenden Jahr 2023 in Kraft treten sollte. Gemäß dem Leiter des Forschungsbereiches „Eingebettete Intelligenz“ am Deutschen Forschungszentrum für Künstliche Intelligenz, Paul Lukowicz, sei eine Regulierung in diesem Bereich dringend notwendig gewesen. Gemäß dem Experten war beispielsweise die Einführung eines Wasserzeichens für von KI-Systemen generierten Inhalten möglich. Auf diese Weise sollten Schäden, beispielsweise durch gezielte oder unbewusste Desinformation vermieden werden. Die Regulierungsmaßnahmen sollten jedoch in einem ausgewogenen Kosten-Nutzen-Verhältnis stehen.

<https://www.tagesschau.de/wirtschaft/technologie/alibaba-ki-chatbot-china-regulierung-101.html>

01.12.23 13:58 Freitag

Nach einer langen Pause in meiner Untersuchung zum Thema künstlicher Intelligenz war es mir nun endlich gelungen, den großen Berg an Arbeit in meiner Firma abzuarbeiten, sodass ich nun wieder mehr Kapazitäten frei hatte, um meine Quellenarbeit fortzusetzen. Mir war bewusst, dass ich nun Quellen bearbeiten musste, die Teils mehr als ein halbes Jahr alt waren. Dies war jedoch nicht so tragisch, da die KI-Entwicklung in der Zwischenzeit offenbar keine weiteren Quantensprünge hingelegt hatte. Dies erschien mir vor dem Hintergrund der geforderten Pause zur KI-Entwicklung nur logisch. Im Bereich der Regulierung von KI-Systemen und -Entwicklern hatte sich jedoch inzwischen einiges getan, ebenso aus wirtschaftlicher Perspektive.

Ethikrat: Künstliche Intelligenz darf menschliche Entfaltung nicht vermindern

Da ich den ursprünglichen Eintrag zum deutschen Ethikrat in dem Sternartikel noch validieren wollte, öffnete ich diesen nun.

Die Einleitung der Stellungnahme des Deutschen Ethikrates vom 20.03.2023 nahm darauf Bezug, dass KI-Systeme bereits in nahezu allen Bereichen des öffentlichen und privaten Lebens Einzug gehalten hatten: von der Medizin, über die Pädagogik, über die Kommunikation im Internet bis hin zu öffentlich-rechtlichen Einrichtungen, wie zum Beispiel im Sozial- und Justizwesen oder der Polizei.

Der Ethikrat stellte die zentrale Forderung auf, dass KI-Anwendungen die menschliche Intelligenz, Verantwortung und Bewertung nicht ersetzen durften und zudem die menschliche Entfaltung erweitern und nicht vermindern sollten. Daher sollten die komplexen Wechselwirkungen zwischen Mensch und Maschine sowie die gesellschaftlichen Effekte beachtet werden.

Im Weiteren untersuchte der Deutsche Ethikrat in seiner Stellungnahme diese Forderungen exemplarisch im Kontext der vier Anwendungsbereiche: Medizin, schulische Bildung, öffentliche Kommunikation und Meinungsbildung sowie in der öffentlichen Verwaltung. Zunächst stellte der Artikel fest, dass keine allgemeingültigen Aussagen zum Einsatz von KI-Systemen gemacht werden konnten und diese immer kontext-, anwendungs- und personenspezifisch erfolgen mussten.

Für den Medizinbereich empfahl der deutsche Ethikrat unter anderem, die Qualitätssicherung bei der Entwicklung und Nutzung von KI-Systemen in den Vordergrund zu stellen, ebenso wie die Vermeidung ärztlicher Kompetenzverluste sowie den Schutz der Privatsphäre von Patientinnen und Patienten. Dieser Schutz sollte im Einklang mit einer intensiven Datennutzung in der medizinischen Forschung in gebracht werden. Das Vertrauensverhältnis zwischen allen beteiligten Personen sollte geschützt werden und „die vollständige Ersetzung medizinischer Fachkräfte“ vermieden werden.

Für den Einsatz von KI-Systemen in der schulischen Bildung empfahl der Ethikrat, dass weiterhin grundlegende Bildungsvorstellungen im Vordergrund stehen sollten und nicht technologische Visionen. Es sollte sich bei dem Einsatz von KI-Systemen lediglich auf Elemente beschränkt werden, die nachweislich die Kompetenzen und die sozialen Interaktionen der Lernenden erweiterten. Zudem sollte Wert auf den Schutz der Privatsphäre und Persönlichkeitsbildung der Schüler gelegt werden.

In der öffentlichen Kommunikation und Meinungsbildung empfahl der Ethikrat die Weiterentwicklung von Regeln für Online-Plattformen hinsichtlich der Auswahl und Moderation von Inhalten sowie zu personalisierter Werbung und zum Datenhandel. Zudem sollte ein besserer Zugang auf Plattformdaten für die Forschung gewährt und die digitale Kommunikationsinfrastruktur gegebenenfalls in die öffentliche Hand gelegt werden.

Beim Einsatz von KI in der öffentlichen Verwaltung riet der Ethikrat zu Ansätzen, die vor Diskriminierung schützten und dem blinden Befolgen maschineller Empfehlungen sollte vorgebeugt werden. Einzelfallbetrachtungen sowie die Einsichts- und Einspruchsrechte von Betroffenen sollten

weiterhin gewährleistet werden. Zudem forderte der Ethikrat, dass bei der Anwendung von KI-Systemen in der Arbeit von Gefahrenabwehrbehörden gesellschaftliche Aushandlungsprozesse über ein angemessenes Verhältnis zwischen Risiken und Chancen solcher Ansätze geführt werden sollten.

In allen Bereichen sollten die KI-Systeme nur zur Entscheidungsunterstützung eingesetzt werden und nicht zur Entscheidungsersetzung. Dennoch konnte es dem Artikel nach zu einer Diffusion der Verantwortung kommen, infolge derer nicht mehr eindeutig geklärt werden konnte, ob die Entscheidung nun von dem Nutzer des KI-Systems oder von dem KI-System selbst getroffen worden ist. Diese Diffusion der Verantwortung galt es zu verhindern. Meiner Einschätzung nach hatte das vor Allem straf- und zivilrechtliche Gründe, aber auch versicherungstechnische Gründe, da bei einem Übergang der Verantwortung auf die KI den Herstellern der KI-Systeme im Schadensfall ihrerseits einen Teil der Verantwortung übertragen werden musste, da die KI-Systeme für die Ergebnisse ihrer Berechnungen nicht haftbar gemacht werden konnten.

Im Weiteren durften menschliche Kontrolloptionen nicht beeinträchtigt werden, Abhängigkeiten von den KI-Systemen, der Missbrauch der Technik sowie unerwünschte Verluste menschlicher Fertigkeiten sollten vermieden werden. Der Datenschutz beziehungsweise die Interessen der Menschen, von denen die Daten stammten sollten im Mittelpunkt stehen und Eingriffe in die Privatsphäre mithilfe effektiver rechtlicher und technischer Vorkehrungen verhindert werden, wobei zugleich eine gemeinwohlorientierte Datennutzung ermöglicht werden sollte.

Zusammenfassend konnte man sagen, dass der deutsche Ethikrat dem Einsatz von KI-Systemen in weiten Teilen des gesellschaftlichen Lebens, ebenso wie in kritischen Bereichen der Exekutive, Legislative, Forschung und Medizin grundsätzlich zustimmte, sofern die von Ihm gegebenen Empfehlungen eingehalten wurden. Demnach sollte der Missbrauch von KI-Systemen vermieden werden, obgleich hierbei eher das gesellschaftliche Wohl, als das „Wohl“ des KI-Systems in Vordergrund stand.

<https://www.ethikrat.org/mitteilungen/mitteilungen/2023/ethikrat-kuenstliche-intelligenz-darf-menschliche-entfaltung-nicht-vermindern/?cookieLevel=not-set>

Innenministerium will KI-Einsatz regulieren

Gemäß dem Artikel von Tagesschau.de vom 16.04.2023 hatte sich das Bundesinnenministerium für einen „klaren Rechtsrahmen“ beim Einsatz von KI-Systemen ausgesprochen. Demnach wurde das Thema „Künstliche Intelligenz“ seit der Vorstellung von ChatGPT intensiv diskutiert. Es sollte eine Balance zwischen Innovationsoffenheit und „Standards für vertrauenswürdige KI“ definiert werden: „Regeln brauche es für grundrechtliche Fragestellungen, datenschutzrechtliche Aspekte aber auch für die Frage, wie Fehler und Diskriminierung vermieden werden könnten.“. Eine Rolle spielten auch die Erklärbarkeit und Nachvollziehbarkeit von KI-Ergebnissen. Man sehe ein großes Potenzial für die Gesellschaft durch die Verwendung von KI-Anwendungen.

Regeln für den Umgang mit KI-Systemen sollten auf nationaler und europäischer Ebene erarbeitet werden. Die SPD-Chefin Saskia Esken sprach sich für eine Kennzeichnungspflicht von KI-Inhalten aus, da das Risiko bestand, dass in Zukunft echte und durch KI generierte, beziehungsweise gefälschte Sachverhalte nicht mehr auseinandergehalten werden konnten.

Der Tagesschau.de Artikel zitierte auch den Digitalminister Volker Wissing von der FDP, der die Fähigkeiten von zeitgenössischen KI-Systemen in der Bild am Sonntag pries und feststellte, dass diese Systeme nun im Alltag angekommen seien und das Leben der Menschen grundlegend veränderten. Daher hatte er die Europäische Union aufgefordert, schnell zu reagieren und künstliche Intelligenz vernünftig zu regulieren, um sicherzustellen, dass die Technologie nur eingesetzt wurde, wenn Sie im Einklang mit europäischen Werten, wie Demokratie, Transparenz und Neutralität stand.

Im Weiteren ging der Tagesschau.de Artikel darauf ein, dass die EU-Staaten bereits im Dezember 2022 erstmals umfassende Regeln für den Einsatz von KI-Systemen festgelegt hatten, um deren Sicherheit zu gewährleisten, sich jedoch noch mit dem Europaparlament auf eine Linie verständigen mussten, bevor die Regeln in Kraft gesetzt werden konnten. Die Sicherheitsanforderungen sollten sich an dem Gefahrenpotenzial der KI-Systeme orientieren und für Regelverstöße hohe Strafen vorgesehen werden.

Abschließend nahm der Artikel noch Bezug auf die Stellungnahme des deutschen Ethikrates und auf die Ankündigung von Chinas Cyberspace Administration, die Entwicklung von künstlicher Intelligenz zu unterstützen. In diesem Zusammenhang wurde der Staatssekretär der ÖVP, Florian Tursky, zitiert, nach dem eine enorme Gefahr von KI-Systemen, welche einer staatlich vorgegebenen Ideologie folgten, für die Demokratien ausgingen, da solche KI-Systeme, wenn sie in Europa auf den Markt kämen beispielsweise den ideologischen Fußabdruck der kommunistischen Partei Chinas hätten.

<https://www.tagesschau.de/inland/ki-regulierung-103.html>

Die Internetseite von ChatGPT

Als Nächstes wollte ich die Internetpräsenz des KI-Systems ChatGPT genauer unter die Lupe nehmen. Diese Betrachtung konnte mir Hinweise über die vorgesehenen Funktionen der Software geben und die Zielstellung der Entwickler bei der Erschaffung des Systems näherbringen, sowie möglicherweise Hinweise auf eine zukünftige Entwicklung des Systems geben. Als ich an diesem 01.12.2023 „ChatGPT“ in die Suchmaschinen eingab, gaben mir ECOSIA.de und google interessanterweise zwei unterschiedliche Seiten als ersten Eintrag in den Suchergebnissen aus. Während ECOSIA mit <https://openai.com/chatgpt> auf die Internetseite des Herstellers der Chatsoftware verwies, verlinkte google direkt auf den Login-Bereich des Tools, welcher jedoch vom Umfang der gegebenen Informationen her vergleichsweise sparsam war: <https://chat.openai.com/auth/login> war in zwei Hauptbereiche unterteilt, die spaltenförmig nebeneinander und ungefähr im Flächenverhältnis 2/3 zu 1/3 angeordnet waren. Die linke Spalte blendete nacheinander stichpunktartig Vorschläge für mögliche Nutzungsmöglichkeiten, wie Gedankenstützen, Optimierung von Nachrichten oder Quellcode, E-Mailentwürfe und diverse private Nutzungsweisen, wie die Recherche von Freizeitaktivitäten, Reiseplanung und Geschenkeauswahl ein. Die rechte Spalte enthielt neben dem Log-In- und Sign-up-Button für die Anmeldung und Registrierung zur Verwendung des Chatbots lediglich einen Link mit der Beschriftung „Privacy policy“, der mutmaßlich auf eine Datenschutzerklärung verweisen sollte, jedoch als Ergebnis nur den Fehler 404 „Die Internetseite existiert nicht.“ hervorbrachte, wie ich pikanterweise feststellen musste. Zudem gab es jedoch einen Verweis auf die Nutzungsbedingungen, deren vollständige Untersuchung ich mir (fahrlässiger Weise) ersparte. Ich stellte aber fest, dass es für die Nutzer aus der europäischen Union, aus der Schweiz und dem Vereinigten Königreich eine eigene unterverlinkte Seite zur Erläuterung der Nutzungsbedingungen gab. Dort war neben der üblichen umfangreichen Erklärung unter Anderem geregelt, wie Nutzer die Software zu verwenden hatten:

„Was Sie nicht tun dürfen. Sie dürfen unsere Dienste nicht für illegale, schädliche oder missbräuchliche Aktivitäten nutzen. Es ist Ihnen zum Beispiel untersagt:

- Unsere Dienste auf eine Art und Weise zu nutzen, die die Rechte anderer verletzt, missbraucht oder gegen sie verstößt.
- Unsere Dienste zu modifizieren, zu kopieren, zu vermieten, zu verkaufen oder zu vertreiben.
- Der Versuch, den Quellcode oder die zugrundeliegenden Komponenten unserer Dienste, einschließlich unserer Modelle, Algorithmen oder Systeme, zurückzuentwickeln, zu

dekompilieren oder herauszufinden, oder jemanden dabei zu unterstützen (es sei denn, diese Einschränkung ist durch geltendes Recht unzulässig).

- Automatisches oder programmgesteuertes Extrahieren von Daten oder Output (wie unten definiert).
- Die Behauptung, dass der Output von Menschen erstellt wurde, obwohl dies nicht der Fall ist.
- Die Beeinträchtigung oder Störung unserer Dienste, einschließlich der Umgehung von Obergrenzen oder -beschränkungen oder der Umgehung von Schutzmaßnahmen oder Sicherheitsvorkehrungen, die wir für unsere Dienste vorgesehen haben.
- Die Verwendung von Output zur Entwicklung von Modellen, die mit OpenAI konkurrieren.“

Zudem wurde in zwei Absätzen die Verantwortlichkeiten für Inhalte festgelegt: „Ihre Inhalte. Sie können Input für die Dienste bereitstellen ("Input") und von den Diensten Output erhalten, die auf dem Input basieren ("Output"). Input und Output werden zusammen als "Inhalte" bezeichnet. Sie sind für den Inhalt verantwortlich, einschließlich der Sicherstellung, dass dieser nicht gegen geltendes Recht oder gegen diese Bedingungen verstößt. Sie sichern zu und gewährleisten, dass Sie über alle Rechte, Lizenzen und Genehmigungen verfügen, die für die Bereitstellung von Input für unsere Dienste erforderlich sind.“

Demnach wurde dem Benutzer der Internetseite die volle Verantwortung für den Output der Software übertragen. Diese wurde dem Nutzer übertragen, obwohl er in den allermeisten Fällen nicht im Vorfeld wissen konnte, mit welchem Output die Software auf einen Input reagieren würde. Diese Regelung hielt ich zumindest für fragwürdig.

„Rechte an den Inhalten. Im Verhältnis zwischen Ihnen und OpenAI und soweit nach geltendem Recht zulässig, behalten Sie (a) Ihre Inhaberrechte am Input und (b) Ihnen stehen die Rechte am Output zu. Wir treten hiermit alle unsere Rechte, Titel und Anteile, falls vorhanden, am Output und in Bezug auf den Output an Sie ab.“

Demnach wurden dem Nutzer durch die Nutzungsbedingungen freundlicherweise die Rechte an dem Output der Software übertragen. Dies war aus meiner Sicht verständlich, da der Nutzen der ausgegebenen Informationen für Menschen, die keine Rechte an dem Output der Software hatten, wahrscheinlich sehr eingeschränkt war, was vermutlich auch zu umfangreichen Zuwiderhandlungen mit den entsprechenden strafrechtlichen Konsequenzen geführt hätte.

Dafür konnte die Software vom Nutzer eingegebene Inhalte weltweit nutzen:

„Unsere Verwendung der Inhalte. Wir können Ihre Inhalte weltweit nutzen, um unsere Dienste bereitzustellen, aufrecht zu erhalten, zu entwickeln und zu verbessern, geltende Gesetze einzuhalten, unsere Bedingungen und Richtlinien durchzusetzen und unsere Dienste sicher zu halten.“

Der Nutzung eingegebener Informationen für das Training der künstlichen Intelligenz konnte man jedoch über die unter <https://privacy.openai.com/policies> angegebene E-Mailadressen widersprechen.

26.01.24

Die Internetseite von OpenAI

Die Internetseite des Softwareanbieters <https://openai.com/chatgpt> ähnelte vom Design her der oben beschriebenen Login-Seite von ChatGPT. Als zusammenfassende Funktionsbeschreibung des Chatbots wurde folgender Satz angegeben: „Get instant answers, find creative inspiration, learn something new.“.

Daneben befanden sich zwei Screenshots der Software, einmal in der Desktopversion und einmal in der mobilen Version für Smartphones, in denen exemplarisch Dialoge mit dem Chatbot dargestellt wurden.

Darunter wurden wiederum eine ganze Reihe verschiedener Themenvorschläge gegeben, in Form von stichpunktartigen Anweisungen, die ChatGPT gegeben werden konnten.

Anmerkung hierzu: Während ich mir diese Übersicht noch einmal anschaute, fiel mir die Analogie von ChatGPT zu der DOS-Eingabe auf. Ursprüngliche DOS-Betriebssysteme arbeiteten ausschließlich auf Basis der Eingabe von Textbefehlen und ähnelten daher vom Grundprinzip her der Eingabe, wie bei ChatGPT, mit dem Unterschied, dass bei ChatGPT sowohl vom Befehlsinhalt, als auch von der Syntax her bei der Eingabe im Prinzip keine Grenzen gesetzt waren. Hier ließ sich noch einmal der fundamentale Unterschied zwischen einem herkömmlichen deterministischen Computerprogramm und einem Machinelearning-zentrierten System erkennen.

Es wurde also beim Prinzip „Eingabe, Verarbeitung, Ausgabe“ wieder ein Stück weit zu den Anfängen der Computertechnik zurückgekehrt, nämlich von der grafischen zurück zu der textbasierten Ein- und Ausgabe. Allerdings ließen sich im Gegensatz zur DOS-Eingabe keine Unterprogramme oder Dateiverwaltungsanweisungen ausführen. Jedoch stellte sich mir die Frage, wie ChatGPT auf solche Anweisungen reagieren würde. Vielleicht sollte ich dies bei Gelegenheit einmal ausprobieren?

Scrollte man die Internetseite von OpenAI weiter herunter, wurden weitere Funktionen des Chatbots aufgeführt. Demnach wurden dem System Funktionen zur Bildverarbeitung und -erzeugung hinzugefügt, ebenso wie die Ein- und Ausgabemöglichkeit von gesprochener Sprache. Diese erheblichen Erweiterungen des Funktionsumfangs warfen natürlich eine ganze Reihe von Möglichkeiten für Anwendungen und von Missbrauch des Systems auf. Sie ließen zudem Rückschlüsse auf eine erhebliche Erweiterung der internen Komplexität und Verarbeitung des Chatbots zu, welche möglicherweise auch durch den Einsatz von Schnittstellen zu anderen (möglicherweise KI-basierten) Systemen umgesetzt worden waren.

Scrollte man die Internetseite weiter herunter, wurden die verschiedenen Zahlungsoptionen, „Free“ bzw. „Plus“ aufgeführt. Demnach gab es die Softwareversion GPT-3.5 für eine kostenlose Nutzung. Sie umfasste den Zugriff aus dem Internet und von den verschiedenen Betriebssystemen und regelmäßige Updates des KI-Modells.

Die Plus-Variante beinhaltete neben den Funktionen der kostenlosen Option die Version GPT-4, der dem Hersteller nach „fähigsten“ Version sowie die oben bereits beschriebenen Funktionen der Bild- und Sprachkommunikation. Zudem wurde die Möglichkeit angeboten, eigene benutzerdefinierte Chatbotvarianten zu erstellen.

Abschließend wurden auf der Seite verschiedene benutzerdefinierte Chatbotvarianten vorgestellt, von denen es gemäß der Beschreibung bereits hunderte funktionspezifische Varianten gibt.

Das Menü von <https://openai.com/chatgpt> enthielt eine Verlinkung zu der Beschreibung des mit dem Chatbot verknüpften Bilderstellungstools DALL-E 3, welches ebenfalls von OpenAI entwickelt worden war. Als Funktionsbeschreibung gab OpenAI folgenden Hinweis: „DALL·E 3 is built natively on ChatGPT, which lets you use ChatGPT as a brainstorming partner and refiner of your prompts. Just ask ChatGPT what you want to see in anything from a simple sentence to a detailed paragraph. When prompted with an idea, ChatGPT will automatically generate tailored, detailed prompts for DALL·E 3 that bring your idea to life. If you like a particular image, but it's not quite right, you can ask ChatGPT to make tweaks with just a few words.“

Im Weiteren gab es auf <https://openai.com/chatgpt> eine Verlinkung zu einer Einführung für Entwickler, welche den ChatBot über API-Schnittstellen ansteuern konnten.

Die kommunizierte Sicherheitsphilosophie von OpenAI

Der Safety-Reiter im Menü verlinkte auf <https://openai.com/safety> und vermittelte einen Eindruck über die Sicherheitsphilosophie des Unternehmens. Demnach sollten KI-Systeme im Einklang mit menschlichen Intentionen und Werten stehen: „AI systems are becoming a part of everyday life. The key is to ensure that these machines are aligned with human intentions and values.“. Im Weiteren wurde das ernstzunehmende Risiko des Missbrauchs von KI-Systemen erkannt und eine sichere und nützliche KI-Entwicklung gemäß der Satzung des Unternehmens priorisiert. Anschließend fand man eine Verknüpfung zu der Satzung des Unternehmens unter <https://openai.com/charter> .

Die Satzung von OpenAI

In der Satzung, die schon bereits seit dem 09.04.2018 bestand und die angewendeten Prinzipien bei der Realisierung von KI-Systemen aufführte, wurde die Strategie des Unternehmens beschrieben. Diese Strategie wurde laut der Internetseite in einem Zeitraum von zwei Jahren unter dem Einfluss des Feedbacks vieler Menschen, sowohl OpenAI-intern, als auch –extern, entwickelt und sollte die Interessen der Menschheit vertreten. Das Potenzial der Systeme, deren Entwicklung OpenAI sich zur Aufgabe gemacht hatte, wurde wie folgt beschrieben: „highly autonomous systems that outperform humans at most economically valuable work“. Im Weiteren wollte OpenAI anderen Parteien helfen, dieses Ziel ebenfalls zu erreichen.

Der Nutzen der zu entwickelnden Systeme wurde betont, ebenso so die Vermeidung der Nutzung der Produkte von OpenAI zum Schaden der Menschheit oder zu unangemessener Konzentration von weltlicher Macht.

Interessenkonflikte der Mitarbeiter des Unternehmens sowie der Teilhaber sollen minimiert werden und in einen Kompromiss mit möglichst breitem Nutzen umgewandelt werden.

Mit Hinblick auf die „Long-term safety“ wurde die Verpflichtung des Unternehmens, Forschungen zur Entwicklung sicherer Systeme zu betreiben und eine breite Anwendung dieser Forschung in der Gesellschaft zu fördern betont.

Das Dilemma des Wettbewerbs im Zusammenhang mit der Entwicklung von KI-Systemen wurde bereits postuliert und eine Verhaltensweise im Zusammenhang damit beschrieben: „We are concerned about late-stage AGI development becoming a competitive race without time for adequate safety precautions. Therefore, if a value-aligned, safety-conscious project comes close to building AGI before we do, we commit to stop competing with and start assisting this project.“

OpenAI strebte gemäß seiner Satzung eine Technologieführerschaft im KI-Sektor an, mit der Begründung, dass das Unternehmen die genannten Richtlinien und Werte nur über diese Technologieführerschaft effektiv in die Gesellschaft einbringen konnte. Sollte jedoch ein anderes Projekt vor OpenAI das angestrebte Ziel erreichen, eine AGI zu erschaffen, würde OpenAI nicht mit diesem Projekt in den Wettbewerb treten, sondern stattdessen damit beginnen, dieses Projekt zukünftig ebenfalls zu unterstützen.

Im letzten Absatz beschrieb OpenAI seinen kooperativen Ansatz. Demnach sollte mit anderen Forschungseinrichtungen und Regelwerks-Institutionen kooperiert werden und eine globale Gemeinschaft kreiert werden, welche an den Herausforderungen der KI-Entwicklung arbeitete. Abschließend wurde darauf eingegangen, dass Sicherheitsbedenken die Veröffentlichung von Forschungsergebnissen des Unternehmens zukünftig voraussichtlich einschränken würden, ausgenommen von Ergebnissen der Sicherheits- Regelwerks- und Grundlagenforschung.

Artificial General Intelligence (AGI)

Bemerkenswerter Weise wurde in der Satzung von OpenAI neben dem Begriff der künstlichen Intelligenz „AI“ ein weiterer Begriff eingeführt, der mir bis dato noch nicht geläufig gewesen war, die AGI (Artificial General Intelligence).

Laut der Wikipedia-Definition auf https://de.wikipedia.org/wiki/Artificial_General_Intelligence war damit ein System gemeint, dass wie folgt charakterisiert wurde: „Artificial General Intelligence (AGI), deutsch Künstliche allgemeine Intelligenz, ist die hypothetische Intelligenz eines Computerprogramms, welches die Fähigkeit besitzt, jede intellektuelle Aufgabe zu verstehen oder zu lernen, die ein Mensch ausführen kann.“

Dieser Punkt warf jedoch einige wesentliche Fragen auf:

Wenn eine Software dazu in der Lage wäre, jede intellektuelle Aufgabe zu verstehen, oder zu erlernen zu der ein Mensch fähig war, wo bliebe denn dann noch ein Unterschied zwischen den „geistigen“ Fähigkeiten eines Menschen und denen einer KI. Anders gesagt: Würde so ein System nicht genau genommen per Definition auch zu Selbstwahrnehmung fähig sein, da es sich bei dieser im Grunde auch nur um eine Form einer intellektuellen Leistung handelte?

Zudem würde die Möglichkeit, jede intellektuelle Aufgabe zu verstehen und zu erlernen in Verbindung mit den Systemen zur Bild- und Tonverarbeitung, die über Schnittstellen an ChatGPT angeschlossen worden waren, bei einer AGI theoretisch auch in gewisser Weise die von mir zuvor bereits beschriebene Kombination von Sensorik und Verarbeitung in Form eines entsprechenden Lernalgorithmusses darstellen, wodurch eine solche AGI wiederum (möglicherweise auch ungewollt) zu einer Selbstwahrnehmung gelangen konnte.

Im Weiteren führte der Eintrag in der Wikipedia die Beschreibung der Künstlichem allgemeinem Intelligenz wie folgt aus: „AGI ist ein Hauptziel von Forschungen zur künstlichen Intelligenz (KI, engl. AI, *artificial intelligence*) und ein häufiges Thema in Science-Fiction und Zukunftsforschung. AGI kann auch als *Strong AI* (starke KI) oder *Full AI* bezeichnet werden – im Gegensatz zu *Strong AI* ist *Weak AI* (schwache KI) nicht dazu gedacht, menschenähnliche kognitive Fähigkeiten und Persönlichkeit zu zeigen, sondern beschränkt sich auf die Verwendung von Software zum Lösen ausgewählter vorbestimmter Probleme oder Argumentationsaufgaben (Expertensystem).“

Stand die Satzung von OpenAI damit im Widerspruch zu dem Pause Giant Letter, oder andersherum? Die Forderung des offenen Briefes, ein KI-System sollte sicher sein, außerhalb eines vernünftigen Zweifels konnte ja auch anders als meine Interpretation verstanden werden, nämlich als Forderung, dass ein solches System unzweifelhaft gewisse definierte Sicherheitsstandards erfüllen sollte, unabhängig davon ob es sich dabei um eine „höhere Intelligenz“ handelte, oder nicht. Demnach stand die Option eine höhere künstliche Intelligenz zu erschaffen möglicherweise durchaus im Raum und hatte der Pause Giant Letter diesem Ziel nur eine Art Moratorium verschafft und nach Ablauf der Frist beziehungsweise, wenn die rechtlichen Grundlagen für KI-Systeme in Form von Gesetzen festgelegt worden waren, würde die Wissenschaft mit der Entwicklung solcher Systeme fortfahren. Wurden die ethischen Komplikationen, die ein solches Vorhaben mit sich brachte also wissentlich in Kauf genommen?

Gemäß dem Wikipedia-Artikel existierten sogar schon zeitliche Prognosen, wann die ersten AGI-Versionen möglicherweise fertiggestellt wurden, unter anderem auch von Sam Altman dem CEO von OpenAI. Die Prognosen hätten zwar eine große zeitliche Bandbreite, es wurde jedoch davon ausgegangen, dass bereits in 2030-Jahren solche Systeme existieren konnten (Stand 07.02.24).

Ich beschloss, den zitierten Wikipedia-Eintrag später noch ein wenig genauer anzuschauen und kehrte zurück zu <https://openai.com/safety>, wo sich direkt unter der Verknüpfung zur Satzung von OpenAI ein weiterer Link befand: „Read our approach to alignment research“, welcher auf die Seite <https://openai.com/blog/our-approach-to-alignment-research> verlinkte.

08.02.24

Die Anpassung von KI-Systemen an menschliche Werte und Ziele

Der oben genannte Hyperlink führte zu einem Artikel vom 24.08.2022, der einen Einblick in die angewendeten Trainingsmethoden für KI-Systeme mit dem Ziel der Anpassung ihrer Reaktionen an menschliche Werte und Ziele gab.

Demnach sollte ein Erfahrungsaufbau der KI-Entwickler durch iterative Versuche zur Anpassung der KI-Systeme generiert werden. Die möglichst fortgeschrittene Anpassung des Verhaltens der KI-Systeme an menschliche Werte war also ein Gegenstand der laufenden Forschung und Dokumentation.

Um den Prozess zu verbessern sollten hierfür wiederum auch eigens KI-Systeme als Assistenten entwickelt werden.

Es wurde in dem Artikel noch einmal auf die Schwierigkeit der Anpassung von hochkompetenten KI-Systemen (bzw. AGIs) hingewiesen, ebenso wie auf das Risiko, welches von solchen AGIs ausging, wenn sie nicht an die menschlichen Werte angepasst worden waren. Daher sollte eine Transparenz über die bei OpenAI angewendeten Methoden zur Anpassung der Systeme hergestellt werden:

„At a high-level, our approach to alignment research focuses on engineering a scalable training signal for very smart AI systems that is aligned with human intent. It has three main pillars:

1. Training AI systems using human feedback
2. Training AI systems to assist human evaluation
3. Training AI systems to do alignment research“

Ein Grundprinzip des Trainingsprozesses wurde hierbei als „Reinforcement Learning“ benannt. Demnach läuft ein Prozess in einer Schleife immer wieder durch und wird durch menschliches Feedback nach und nach angepasst, bis das gewünschte Verhalten des KI-Systems beispielsweise in Bezug auf die Anpassung an menschliche Werte und Ziele möglichst optimal erreicht wird. Als Trainingsziele wurden explizit und implizit übermittelte Werte, wie beispielsweise Ehrlichkeit, Fairness und Sicherheit angegeben.

Unterschiedliche Versionen von GPT zeigten eine unterschiedliche Anpassung nach diesen Kriterien, je nachdem wie viel diese trainiert worden waren und welche Trainingsmethoden angewendet worden waren.

Es waren KI-Modelle entwickelt worden, die eine besonders gute Anpassung an menschliche Werte und Ziele zeigten. Jedoch wurden auch hier noch Probleme bei der Anpassung identifiziert: „Yet today’s versions of InstructGPT are quite far from fully aligned: they sometimes fail to follow simple instructions, aren’t always truthful, don’t reliably refuse harmful tasks, and sometimes give biased or toxic responses.“ Auch wurde bei solchen Systemen ein Verlust von Kreativität im Vergleich zu Vorgängerversionen beschrieben.

Zudem wurde festgestellt, dass ein Training auf Basis von menschlichem Feedback allein nicht hinreichend sein konnte um AGIs an die menschlichen Werte und Ziele optimal anzupassen. Dies lag unter anderem in der Begrenzung menschlicher Fähigkeiten zur Evaluation der Ergebnisse welche solche Systeme unter anderem generieren sollten (z.B. die Zusammenfassung von Büchern oder die

Suche nach Fehlern in Quelltexten oder wissenschaftlichen Arbeiten). In dem Artikel ist zudem schon zuvor auf die Herausforderung hingewiesen worden, festzulegen an welche Menschen die KI-Systeme nun tatsächlich angepasst werden sollten.

Daher sollten wiederum KI-Systeme eingesetzt werden, um Menschen bei dem Training komplexerer KI-Systeme, deren Aufgaben bzw. Ergebnisse zu schwierig waren, um auf einfache Weise durch Menschen evaluiert zu werden zu assistieren. Verschiedene Verknüpfungen zu Artikeln mit Forschungsergebnissen wurden nun aufgezählt: recursive reward modeling, debate und iterated amplification, auf welche ich hier nicht im Detail eingehen wollte.

So konnte ein Mensch beispielsweise eine computergenerierte Zusammenfassung von Büchern besser einschätzen auf Basis von ebenfalls durch einen Computer generierten Zusammenfassungen einzelner Artikel eines Buches.

Ein anderes genanntes Beispiel war die Verwendung von Suchmaschinen, welche das Internet nach Hyperlinks und Artikeln zu bestimmen KI-generierten „Fakten“, die ein Mensch evaluieren sollte, durchsuchten und diese Informationen dem Menschen zur Verfügung stellten, oder sogar auf Basis der dort gegebenen Informationen selbst eine Evaluation der „Fakten“ durchführen konnten.

Auch wurde dem OpenAI-Artikel zufolge ein KI-Modell entwickelt, mit dem Ziel, seine eigenen Ausgaben zu evaluieren, um damit Menschen zu unterstützen. Die Effektivität der auf diese Weise unterstützten Menschen bei der Suche nach Fehlern in KI-Ausgaben sei demnach um 50% höher gewesen, als ohne den KI-Support.

Der Abschnitt zum Einsatz von KI-Assistenten bei dem Training von künstlicher Intelligenz endete mit der Feststellung, dass diese Assistenzsysteme auch bei der Evaluation von sehr kreativen, irreführenden oder gar betrügerischen Outputs der zu trainieren (höheren) KI-Systeme funktionieren mussten und dass dieses Ziel nur durch Lerneffekte über die Entwicklung weiterer Systeme erreicht werden konnte.

Der nächste Abschnitt „Training AI systems to do alignment research“ befasste sich nun mit noch eigenständigeren Systemen. Demnach sollten KI-Systeme Forschungsaufgaben selbst übernehmen. Es konnte jedoch keine „unendliche“ Verselbstständigung von KI-Systemen bei der Forschung zum Ziel der Anpassung an menschliche Werte und Ziele geben und es wurde erwartet, dass mit fortschreitender Entwicklung der KI-Systeme neue, bisher unvorhergesehene Probleme bei der Anpassung auftreten würden.

Daher sollten die KI-Systeme die Arbeit in Kooperation mit Menschen machen, dabei jedoch zunehmend mehr von der Forschungsarbeit selbst übernehmen und schließlich weitestgehend selbst, unter der Aufsicht von Menschen die Forschungsarbeiten durchführen. Die Menschen sollten wiederum bei der Evaluation der Forschungsergebnisse Unterstützung von KI-Systemen bekommen.

Auch hier wurde wieder darauf hingewiesen, dass für die Forschung prädestinierte KI-Sprachmodelle Zugriff auf die Informationsmengen des Internets haben sollten, da diese, mit bestimmten Filtern versehen, im Großen und Ganzen die Werte der Menschen repräsentierten.

Der Artikel beendete den Abschnitt, dass zum Zeitpunkt seiner Erstellung noch kein hinreichend fortgeschrittenes System existierte, um einen wesentlichen Beitrag zu der „Alignment“-Forschung zu leisten, dass der Zugriff der Öffentlichkeit zu einem solchen System jedoch gewährt werden würde, sobald dieses entwickelt worden war.

Abschließend führte der Artikel noch einige mögliche Grenzen oder Probleme auf, die bei dem (KI-gestützten) Training von KI-Systemen zur Anpassung an menschliche Werte und Ziele auftreten

konnten, die z.B. die Robustheit der Systeme betrafen, aber auch in Bezug auf Inkonsistenzen bei den zum Training eingesetzten KI-Assistenten, oder wenn aufgrund unerwarteter Paradigmenwechsel die bisherigen Trainingsmodelle komplett in Frage gestellt werden mussten und nicht zuletzt wurde auch „höheren“ KI-Assistenten zum Training von AGIs beispielsweise ein gewisses Risiko zugesprochen sofern diese nicht selbst hinreichend an menschliche Ziele und Werten angepasst worden waren. Dem unbenommen sollte die Forschung weiter fortgesetzt werden und abschließend wurde daher auf die Recruiting-Seite von OpenAI verwiesen.

09.02.24

Ich kehrte zurück zu der Internetseite <https://openai.com/safety>, um mir die weitere Kommunikation des Unternehmens zum Thema KI-Sicherheit anzuschauen. Der nächste Abschnitt erläuterte die personelle Aufstellung des Unternehmens im Hinblick auf den Sicherheitsaspekt bei der Entwicklung von künstlichen Intelligenzen. Demnach hatte das Unternehmen drei verschiedene Teams aufgestellt, um ein möglichst breites Feld von Sicherheitsrisiken abzudecken.

Die Mitarbeiter-Teams von OpenAI mit dem Fokus auf der Sicherheit

Das „Safety Systems team“ hatte seinen Fokus bei den Risiken, die bei der Bereitstellung der aktuellen KI-Software, wie beispielsweise von ChatGPT an Kunden entstanden, das „Superalignment team“ fokussierte sich auf die Anpassung von „Superintelligenzen“, das „Preparedness team“ betrachtete demnach die zukünftigen bzw. sich zu aktuell in der Entwicklung befindlichen KI-Systeme, welche als Zwischenstufe zu den „Superintelligenzen“ betrachtet werden konnten, deren Entwicklung das Unternehmen sich zum Ziel gesetzt hatte. Die einzelnen Teams kooperierten untereinander und mit der Regelwerksabteilung des Unternehmens.

Die Verlinkung auf die Seite des Safety Systems Teams führte zu einem Artikel, in dem der Ansatz und die von dem Team zu bearbeitenden Probleme bei der Bereitstellung von KI-Systemen für Kunden bzw. die Gesellschaft beschrieben wurden. Demnach setzte der Einsatz von zukünftig entwickelten AGI-Systemen das Training und den Einsatz gegenwärtiger KI-Systeme unter Feldbedingungen voraus, um das sichere und ethische Verhalten der KI-Systeme unter solchen Bedingungen zu gewährleisten, eine sichere Infrastruktur zu entwickeln und Erfahrungen zu sammeln. Als Problemstellung wurde unter anderem folgende Frage gestellt:

„How do we ensure our models robustly avoid giving unsafe or inappropriate answers, while also still giving useful and trustworthy answers in a wide range of applications, from high-stakes domains to playful applications?“

Hierbei war es demnach schwierig, eine exakte Grenze zu ziehen, um eine Überreglementierung des KI-Verhaltens zu verhindern, da sonst möglicherweise der Nutzen des Systems eingeschränkt werden würde.

Die Implementierung von Sicherheitssystemen, die Alignment-Forschung und die Durchführung von Risikoanalysen, wie bereits zuvor beschrieben, wurden betrieben, ebenso wie die Erstellung von Regelwerken für die Interaktion zwischen Menschen und KI-Systemen. Schließlich sollte auch direktes Feedback der Mitarbeiter an die KI-Systeme, deren Anpassung an menschliche Werte und Ziele schärfen.

Das Preparedness Team sollte hypothetische Risiken in messbare Werte umwandeln und datengetriebene Vorhersagen dazu erstellen. Auf die Ergebnisse wurden dann Risikoanalysen angewendet. Es waren Gefahrenkategorien definiert worden: „Cybersecurity“, „CBRN (chemical, biological, radiological, nuclear threats)“, „persuasion“ und „model autonomy“. Für die einzelnen Kategorien sollten Schwellwerte für Risiken definiert werden, anhand derer die Modelle in Risikokategorien eingeteilt werden konnten „Low“, „Medium“, „High“ und „Critical“.

KI-Modelle, deren Risiko als „Medium“ oder niedriger eingestuft worden waren, wurden für die Bereitstellung an Kunden zugelassen. KI-Modelle, deren Risiko „High“ oder niedriger eingestuft worden waren, wurden zur Weiterentwicklung zugelassen.

Ein funktionsübergreifendes Team sollte die Ergebnisse der Risikoanalysen überwachen, welche sich bis in die Führungsetagen des Unternehmens erstreckte, um auf der Grundlage der Analysen wesentliche Entscheidungen treffen zu können. Zudem sollten externe Audits die Arbeit der Sicherheitsabteilungen zusätzlich absichern und wiederum Kooperationen mit externen Partnern bei der Bekämpfung von Risiken und von Missbrauch solcher KI-Systemen helfen.

Der Verweis von der „Safety“-Seite von OpenAI auf die Beschreibung der Arbeit des Superalignment teams führte auf die Internetseite <https://openai.com/blog/introducing-superalignment>. Der Artikel begann mit der Erklärung, dass es wissenschaftliche und technische Fortschritte erforderte, um KI-Systeme zu steuern und zu kontrollieren, welche sehr viel intelligenter sein würden, als Menschen. Dem eigens hierfür gegründete Forschungsteam wurden 20% der dem Unternehmen zur Verfügung stehenden Rechenleistung zur Verfügung gestellt, mit dem Ziel, die Problemlösung in einem Zeitraum von vier Jahren zu anzuheben.

„Superintelligence will be the most impactful technology humanity has ever invented, and could help us solve many of the world’s most important problems. But the vast power of superintelligence could also be very dangerous, and could lead to the disempowerment of humanity or even human extinction.“

Es wurde der Glaube geäußert, dass eine KI-Superintelligenz bereits in den zweitausendzwanziger bis -dreißiger Jahren entwickelt werden würde und zudem wurde noch eine begriffliche Trennung zwischen einer Superintelligenz und der AGI gemacht. Demnach war eine KI-Superintelligenz eine noch höher entwickelte Technologie, als die AGI.

Daher wurde hier als grundlegende Problemstellung die Frage formuliert:

„How do we ensure AI systems much smarter than humans follow human intent?“

Laut OpenAI gab es gegenwärtig keine Lösung für dieses Problem. Es wurde davon ausgegangen, dass Menschen möglicherweise nicht in der Lage waren, zuverlässig solche Systeme zu überwachen, daher sollte die technologische Ansatz bei der Lösung des Problems helfen.

Dafür sollte ein KI-System eingesetzt werden, dessen Intelligenz ungefähr auf menschlichem Niveau war. Dieses System sollte dann mit einer großen Rechenleistung ausgestattet werden, um im Rahmen von „automatischer Forschung“ in iterativen Schritten eine etwaige Superintelligenz anpassen zu können.

Die generelle Vorgehensweise wurde mit den folgenden Schritten definiert:

1. „To provide a training signal on tasks that are difficult for humans to evaluate, we can leverage AI systems to assist evaluation of other AI systems (*scalable oversight*). In addition, we want to understand and control how our models generalize our oversight to tasks we can’t supervise (*generalization*).“
2. To validate the alignment of our systems, we automate search for problematic behavior (*robustness*) and problematic internals (*automated interpretability*).“
3. Finally, we can test our entire pipeline by deliberately training misaligned models, and confirming that our techniques detect the worst kinds of misalignments (*adversarial testing*).“

Im Ergebnis erhoffte man sich für die Zukunft eine Entwicklung weiterer Forschungsprioritäten.

Im Weiteren erhoffe man sich, die o.g. Zielstellung zu erreichen, superintelligente Systeme an menschliche Werte und Ziele anzupassen, war sich aber bewusst, dass zu der Lösung des Problems auch gehören würde, dass man überzeugende Beweise und Argumente dafür fand, dass das Problem tatsächlich gelöst worden war.

Um dieses Ziel im Weiteren zu erreichen, sollten machine-learning-Experten rekrutiert werden, um das Problem gemeinsam mit Top-Wissenschaftlern von OpenAI anzugehen. Die Forschungsergebnisse sollten veröffentlicht werden.

Abschließend wurden in dem Artikel noch einmal einige allgemeine Risiken, die bei dem Einsatz von KI-Systemen, bzw. deren öffentlicher Bereitstellung auftreten konnten aufgeführt: Missbrauch, ökonomische Verzerrungen, Desinformation, Verstärkung von Vorurteilen oder Diskriminierung, menschliche Abhängigkeit von künstlicher Intelligenz, oder blindes Vertrauen von Menschen in KI-Systeme.

Auf der Safety-Seite <https://openai.com/safety> von OpenAI wurden neben den o.g. Artikeln weitere Blogartikel des Unternehmens mit Forschungsergebnissen zu dem Thema aufgeführt, von deren Betrachtung ich nun absah, da diese hier den Rahmen gesprengt hätte. Zudem wurde noch einmal darauf hingewiesen, dass mit anderen in der Industrie führenden Unternehmen und mit externen Regelwerkern kooperiert wurde, um eine vertrauenswürdige Weiterentwicklung von KI-Systemen zu garantieren.

Abschließend wurde hier noch ein Hyperlink aufgeführt in dem die Prinzipien des Unternehmens mit Blick auf die KI-Entwicklung und Bereitstellung von KI-Systemen aufgeführt worden sind (<https://openai.com/safety-standards>):

„Minimize harm

We will build safety into our AI tools where possible, and work hard to aggressively reduce harms posed by the misuse or abuse of our AI tools.

Build trust

Alongside our user and developer community, we'll share the responsibility of supporting safe, beneficial applications of our technology.

Learn and iterate

We will observe and analyze how our models behave and are used and seek input on our approach to safety in order to improve our systems over time.

Be a pioneer in trust and safety

We will support research into the unique trust and safety challenges posed by generative AI, to help improve safety beyond our ecosystem.“

Last but not least gab es im Menü der Internetseite von OpenAI noch einen „Company“-Reiter, auf den ich nicht weiter im Detail eingehen wollte, der jedoch zum einen auf eine Beschreibung des Unternehmens verlinkte, wie auch auf die Bereits zuvor von mir erfasste Satzung und im Weiteren einen Überblick über die Produkte und Erfolge, die das Unternehmen bereits zu verzeichnen hatte gab.

Ein kurzes Fazit zur Internetseite von OpenAI zum Thema ChatGPT

Ich wollte nun einmal feststellen, dass zwar ethische Aspekte von OpenAI nicht primär auf der Internetseite in den Fokus gestellt worden waren, das Thema der Sicherheit von künstlichen Intelligenzen aber durchaus einen großen Stellenwert für das Unternehmen einzunehmen schien. Auch erklärte sich das von mir beobachtete, scheinbar menschenähnliche Verhalten von ChatGPT vor dem Hintergrund, dass KI-Systeme offenbar mit großem Aufwand darauf trainiert wurden, eben ein solches Verhalten an den Tag zu legen, nämlich die Anpassung an menschliche Ziele und Werte.

13.02.24

Der Wikipediaeintrag zu OpenAI

Gemäß <https://de.wikipedia.org/wiki/OpenAI> war OpenAI im Jahr 2015 gegründet worden, um Risiken, wie der Verdrängung der menschlichen Spezies durch eine superintelligente künstliche Intelligenz bei der Entwicklung künstlicher Intelligenz entgegenzutreten.

Daher wurde das Unternehmen als Open-Source-Non-Profit-Organisation konzipiert. Damit sollte die Unabhängigkeit der Organisation gegenüber Geldgebern und deren Interessen sichergestellt werden. Für die Forschung sollte der Freiraum geschaffen werden, sich auf lange Sicht positiv auf die Gesellschaft auswirken zu können.^[4] Die Organisation wollte eine „freie Zusammenarbeit“ mit anderen Institutionen^[5] und Forschern ermöglichen, indem sie ihre Patente und Forschungsergebnisse für die Öffentlichkeit zugänglich machte.

Gemäß dem Wikipediaeintrag profitierte OpenAI anfangs so sehr von seinen guten Absichten und dem Image, dass es trotz schlechterer Gehälter im Vergleich zu anderen KI-Entwicklern wie Google oder Facebook zahlreiche Top-Entwickler für sich gewinnen konnte.

Nachdem die bis 2019 erworbenen finanziellen Mittel von etwa 130 Millionen US-Dollar für die Weiterentwicklung der Organisation nicht mehr ausreichten, wurde die gewinnorientierte Tochterfirma OpenAI Global, LLC gegründet, deren Investoren begrenzte Erträge erhielten (bis zu dem einhundertfachen der ursprünglichen Investition). Die Nonprofit-Organisation Open AI, Inc., übte demnach weiterhin die Kontrolle über diese Tochterfirma aus und erhielt deren allfällig überschüssige Erträge.

Größter Investor von OpenAI war Microsoft^[3], welcher bis zum Jahre 2023 zwölf Milliarden US-Dollar in die Organisation gesteckt hatte und OpenAI sein Cloud-Computing-System und Supercomputer zur Verfügung gestellt hatte, damit OpenAI seine eigenen Programme betreiben konnte, bzw. um die durch die von OpenAI betriebenen Programme anfallenden riesigen Datenmengen zu verarbeiten sowie die eigene Forschung vorantreiben zu können.^{[16][17]} Im November 2022 war der KI-Chatbot ChatGPT 3.5 freigeschaltet worden und binnen weniger Tage nutzten mehr als eine Millionen User weltweit das Programm^[17].

Innerhalb von 2 Monaten hätte das Unternehmen gemäß techcrunch.com mehr als 100 Millionen Nutzer weltweit. ChatGPT war demnach die am schnellsten wachsende Plattform aller Zeiten^[18]. Im Herbst 2023 wurde OpenAI mit einem Firmenwert von 80 bis 90 Milliarden Dollar bewertet^[19].

Gemäß einem in der Wikipedia zitierten Artikel von www.wired.com war der Ansatz, das Risiko der Schadensverursachung durch künstliche Intelligenz zu reduzieren, in dem man diese möglichst vielen Menschen zur Verfügung stellte, in der Community umstritten. Elon Musk der gemäß dem Wikipediaartikel diese Strategie mit Sam Altman ursprünglich erdacht habe, damit nicht ein kleiner Teil der Menschheit die Vorteile der KI allein für sich beanspruchen konnte, hingegen kritisierte im Jahr 2023 den profitorientierten Ansatz^[20]. Jüngere KI-Modelle der Organisation waren rechtlich geschützt worden und wurden nicht mehr offen dokumentiert. Daher sei gemäß Wikipedia die Frage nach einer unlauteren Machtkonzentration und fehlender Transparenz aufgekommen^{[21][22]}.

Die Begründung von OpenAI für diese Vorgehensweise ging aus einem von Netzpolitik.org^[21] zitierten Forschungsdokument der Organisation hervor: „Angesichts sowohl der Konkurrenzsituation und der sicherheitstechnischen Implikationen von großen Modellen wie GPT-4, enthält dieser Bericht keine weiteren Details über die Architektur, die Hardware, den Aufbau der Trainingsdaten, Datensatzkonstruktion, Trainingsmethode oder Ähnliches.“

[3] <https://www.theverge.com/2023/11/20/23967515/sam-altman-openai-board-fired-new-ceo>

[4] <https://www.bbc.com/news/technology-35082344>

[5] <https://www.popsci.com/new-openai-artificial-intelligence-group-formed-by-elon-musk-peter-thiel-and-more/>

[16] <https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai>

[17] <https://www.cnbc.com/2023/01/23/microsoft-announces-multibillion-dollar-investment-in-chatgpt-maker-openai.html>

[18] <https://techcrunch.com/2023/11/06/openais-chatgpt-now-has-100-million-weekly-active-users/>

[19] <https://www.handelsblatt.com/technik/ki/chatgpt-openai-koennte-bald-zu-den-teuersten-start-ups-weltweit-gehoren/29413910.html>

[21] <https://netzpolitik.org/2023/ki-forschung-openai-ist-jetzt-closedai/>

[22] <https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview>

Die Wikipedia-Seite zum GPT-Modell (Generative pre-trained transformers)

https://de.wikipedia.org/wiki/Generativer_vortrainierter_Transformer

enthielt unter anderem interessante Angaben zur eingesetzten Datengrundlage, zur aufgewendeten Rechenleistung sowie zu spezifischen Modifikationen, die auf die einzelnen GPT-Versionen von OpenAI angewendet worden waren.

Demnach waren im Trainingsprozess von „GPT-3“ 175 Milliarden Parameter berechnet worden und die dem Training zur Grunde gelegte Datenbasis bestand aus 570 Gigabyte WebText, der englischsprachigen Wikipedia und zwei Sammlungen von Büchern. Die aufgewendete Rechenleistung für das Training der KI betrug demnach $3,1^{23}$ FLOP.

Die Angabe „Parameter“ bezog sich dabei auf die Gewichtung zwischen Eingabe i und Neuron j mit denen in den verwendeten künstlichen neuronalen Netzwerken die Ausgabe von Antworten der GPTs in Abhängigkeit von der Eingabe berechnet wurde. ^[x]

[x] [\[https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz#Allgemeine_Probleme\]](https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz#Allgemeine_Probleme)

14.02.24

Da die Auftragslage bei mir in der Firma vorübergehend etwas abgenommen hatte, hatte ich an diesem Tag, wie an den Tagen zuvor, nach der erledigten Arbeit im Büro noch ein wenig Zeit gefunden, an meinen Recherchen zum KI-Thema weiterzuarbeiten.

Als mir am Vorabend die Zeit zur Recherche in der Wikipedia ausging, beendete ich meinen kleinen virtuellen Ausflug, um den Abend ruhig anzugehen. Das Thema künstliche Intelligenz hatte mich jedoch wieder in seinen Bann gezogen und so entwickelte sich in meinem Kopf bereits ein Entwurf für das nächste Kapitel, den ich zunächst in meinem Handy festhielt, bevor ich schlafen ging.

Eine Bewertung meiner bisherigen Rechercheergebnisse zum KI-Thema

Nach meiner Konsultation der oben genannten Quellen kam ich mir ein wenig naiv und „von gestern“ vor, da ich vor nicht einmal einem Jahr, nachdem ich meine erste Erfahrung mit ChatGPT gemacht hatte, aufgrund der Fortschrittlichkeit des Systems so aus dem Häuschen gewesen war, dass ich direkt versucht hatte, die Medien darüber zu informieren. Damals wusste ich nicht, dass der Hype um ChatGPT bereits seit vier Monaten anhielt und über hundert Millionen Menschen erreicht hatte. Ich war ein wenig auf die Illusion hereingefallen, die mit der Software einherging und hoffte, oder fürchtete, dass ich hier auf etwas gestoßen war, von dem außer mir scheinbar niemand etwas wusste: eine sehr weit fortgeschrittene künstliche Intelligenz.

Die Frage war nun, wieso ich bis zu jenem Märztag noch nichts davon mitbekommen hatte, obwohl ich mich beinahe täglich über das Weltgeschehen in den Nachrichten informierte? Hatte Tagesschau.de, meine zuverlässigste, bisweilen jedoch einzige Quelle über den digitalen Durchbruch nicht berichtet oder war mir der Artikel einfach nur durch die Lappen gegangen? Ich vermutete, dass Tagesschau.de als seriöses Nachrichtenportal nicht, oder nur in seltenen Fällen über die Produkte und Erfolge von Privatunternehmen berichtete, da solche Themen nicht in den Schwerpunkt eines solchen Nachrichtenportals fielen und das die Internetseite vor allem neutral solchen Privatunternehmen gegenüberstehen musste. Dies lag meiner Ansicht nach unter anderem darin begründet, dass der Fortschritt einer Technologie oder eines spezifischen Unternehmens mit hoher Wahrscheinlichkeit Einschnitte für andere Unternehmen und deren eingesetzte Technologien zur Folge haben würde und somit war eine Art von unbezahlter Werbung für ein Privatunternehmen bei öffentlich-rechtlichen Medien auszuschließen.

So hatte ChatGPT mit ziemlicher Wahrscheinlichkeit negative Auswirkungen auf die Besucherzahlen der Wikipedia. Wikipedia und die Onlinemedien hatten seinerzeit wiederum zu einem Umsatzeinbruch bei den Printmedien geführt, ähnlich wie die Einführung von Streamingdiensten zu einem Verschwinden von den meisten Videotheken geführt hatte. Die Erfindung des Buchdruckes galt im Allgemeinen als einer der wichtigsten technologischen Meilensteine in der Geschichte der Menschheit. Waren aber die flächendeckende Einführung des Internets und die damit einhergehende Verbreitung von Wissen und schließlich die darauf aufbauende allgemeine Verfügbarkeit von künstlicher Intelligenz nicht ähnlich große Meilensteine in unserer kulturellen Evolution?

Wir lebten schon in einer seltsamen Zeit. Unsere Gesellschaft übernahm solche technologischen Quantensprünge scheinbar einmütig und mit einer solchen stoischen Gelassenheit, dass man leicht einen verpassen konnte, wenn man nicht aufpasste. Andererseits vermutete ich, dass ein großer Teil der psychischen Erkrankungen von modernen Menschen ihre Wurzeln in eben dieser Schnelllebigkeit unserer Zeit hatte. Würde mein Buch nach seiner Fertigstellung noch irgendeine Relevanz haben? Es lag zum gegenwärtigen Zeitpunkt durchaus im Rahmen des Möglichen, dass in nicht allzu ferner Zukunft KI-Systeme den Menschen ihre automatisch generierten und auf den Hörer zugeschnittenen Arbeiten mit großer Eloquenz und Pathos vorlasen.

Die Frage war auch: sollte ich mich als unabhängiger Beobachter auf die Seite der Befürworter von höher entwickelten KI-Systemen schlagen und meinen Nutzen aus dieser Entwicklung ziehen? Wahrscheinlich würde ich mich ein wenig unglaublich machen, wenn ich als eine Art Untergangsprediger auftreten würde, während scheinbar die halbe akademische IT-Gesellschaft mit hohem Elan daran arbeitete, ein solches KI-System zu erschaffen. Wissen an sich war ja per se erstmal

etwas Gutes. Ich hatte mal einen Spruch gehört, nach dem Wissen nur denjenigen schadete, die es nicht hatten. Ich hatte mir dazu ein Gleichnis ausgedacht: Wissen war so etwas wie ein Haufen Holzspäne. Holzspäne konnte man von der Menge her beliebig proportionieren. Man konnte die Späne vielseitig einsetzen, zum Pressen von Holzplatten beispielsweise, um daraus Möbel oder Fertighäuser zu bauen, oder man nutzte sie als einen hervorragenden Treibstoff für Holzöfen oder für viele andere Einsatzzwecke. Holzspäne waren also auch etwas Gutes. Häufte man jedoch zu viel davon an und hielt sich nicht an die erforderlichen Sicherheitsmaßnahmen, konnte ein Funke den ganzen Haufen in Brand setzen und damit ein Inferno verursachen.

Wissen hatte vielfältige Einsatzzwecke und war von einem unbestreitbaren Nutzen für unsere Gesellschaft. Eine künstliche Intelligenz wurde genutzt zur Verbreitung und Vermehrung von Wissen und war demnach einfach ein Werkzeug, um dieses hohe Gut zu befördern. Und wie im Allgemeinen bekannt war, konnte jedes Werkzeug entweder als Werkzeug oder als Waffe genutzt werden. Aber hörte man auf, Schraubendreher oder Rohrzangen zu produzieren, nur, weil sie theoretisch von den meisten Menschen einfach in eine Hieb- oder Stichwaffe umgewandelt werden konnten? Natürlich nicht. Die Baumärkte waren voll von solchen und ähnlichen Werkzeugen, denn schließlich hatten diese zum Aufbau unserer modernen Gesellschaft einen großen Beitrag geleistet.

Ich beschloss also, die Entwicklung weiter zu beobachten, jedoch nicht vollständig auf den Einsatz der Chatbots zu verzichten, deren Nutzen mit zunehmender Entwicklung wahrscheinlich sogar noch zunehmen würde. Das ChatGPT schon jetzt einen Beitrag zur Wissensvermehrung in der Bevölkerung leistete, konnte bei den hohen Nutzerzahlen wohl kaum bestritten werden. Eine Verbreitung fehlerhafter Informationen wäre bei so vielen Menschen in einer freien Gesellschaft mit ziemlicher Sicherheit jedenfalls aufgefallen, dachte ich.

Welche Auswirkungen würde diese Entwicklung nun auf Wikipedia und co. haben? Würde sich die Pflege der open-source Wissensbibliothek in der Zukunft noch rechnen, wenn die Nutzerzahlen demnächst abnehmen? Ich beschloss, dem Portal, dem ich so viel zu verdanken hatte, demnächst wieder einmal einen kleinen Geldbetrag zu spenden, wenn dieses wieder einen seiner Spendenaufrufe startete. Welche Gegenleistung hatte Wikipedia von OpenAI für die Verwendung seiner Informationen bekommen? Welche Gegenleistung hatten die zahlreichen, ungefragt gecrawlten Internetseiten für Ihre Verwendung als Datenbasis zur Berechnung des KI-Modells bekommen? Welche Gegenleistung hatten die Buchautoren bekommen? Ich wusste aus dem Wikipedia Artikel, dass zu einigen dieser Fragen aktuell Rechtsprozesse liefen und in anderen Fällen auch schon eine Einigung gefunden worden war.

Nun musste ich auch eine weitere Vorgehensweise für meine Arbeit zu diesem Thema und einen Umgang mit meinem Buch finden. Während der letzten Monate war ich ja vorwiegend mit anderen Dingen beschäftigt gewesen, wie der Arbeit in meiner Firma, oder dem Nachgehen meiner anderen persönlichen Interessen. Obgleich ich nebenher nur die allgemeinen Medien frequentiert hatte, waren aufgrund der rasanten technologischen und gesellschaftlichen Entwicklungen im Bereich der künstlichen Intelligenz eine Reihe von Artikeln zu meiner Linksammlung hinzugekommen. Ein Umstand, der durch meine Recherchetätigkeit nicht eben besser wurde. Zudem wurden, wie ich nun wusste von OpenAI Forschungsberichte zu dem Thema veröffentlicht: <https://cdn.openai.com/papers/gpt-4.pdf>

Folgerichtig war ich mit der Untersuchung der Entwicklung weiter ins Hintertreffen geraten.

Andererseits machte es keinen Sinn, die riesigen Informationsmengen der Wikipedia oder des OpenAI-Blogs zum Thema „Künstliche Intelligenz“ in diesem Buch zu verarbeiten. Die Einträge zum Thema des GPT-Modells auf dem der Chatbot basierte, zur AGI (Artificial General Intelligence) oder zu neuronalen Netzwerken waren zwar sicher sehr interessant, waren aber sowieso allgemein zugänglich. Zudem

wollte ich die technologische Betrachtung der künstlichen Intelligenz in diesem Buch ja weitgehend ausklammern und mich auf die ethischen Aspekte konzentrieren. Hochwertige Fachbücher zum Thema technologische KI-Entwicklung gab es ja auch genug.

Ich beschloss also, die Wikipedia auf sich beruhen zu lassen und aus meiner Linksammlung eine kleine, aber relevante Auswahl mit Artikeln zu treffen, welche ich im Rahmen meiner Quellenrecherche noch untersuchen wollte. Anschließend konnte ich dem ChatBot noch einige Fragen stellen und diese dann bewerten, bevor ich das Buch möglichst schnell zu Ende brachte, bevor ich mit meiner Arbeit wieder den Anschluss verlor.

19.02.24

Bevor ich eine Auswahl der relevantesten Quellen treffen konnte, musste ich mir erstmal einen Überblick über meine Quellensammlung verschaffen und herausfinden, was in der letzten Zeit an relevanten Themen hinzugekommen ist. Aus meinen Notizen konnte ich Themen entnehmen, wie „ChatGPT Hack“ und „Schnittstellen für ChatGPT“. Allerdings konnte ich die den Themen zugrundeliegenden Artikel, die ich vor mehr als einem Dreivierteljahr überflogen hatte, nicht mehr ausfindig machen. Stattdessen gab es bei google neue, aktuellere Beiträge. Ich ordnete diese Beiträge in meine Quellensammlung ein und las ein paar neue Artikel, unter anderem einen Jahresrückblick für 2023 von ZDF.de zum Thema „Künstliche Intelligenz“ vom 28.12.2023.

Die Zeit reichte letztendlich nicht aus, um eine Aufstellung der von mir zu behandelnden Themen zu machen. Ein Punkt ist mir bei der Recherche jedoch offenbar geworden. Wenn ich das Thema „Künstliche Intelligenz“ weiter behandeln wollte, musste ich erst einmal versuchen ungefähr zu erfassen, was unter dem Begriff „Intelligenz“ zu verstehen war. Dies war vermutlich ein Fass ohne Boden, aber ein gewisses Grundverständnis des Begriffs war als Basis für die weitere Behandlung des Themas dem Anschein nach unerlässlich. Daher wollte ich hier zumindest den Wikipedia-Eintrag zu dem Thema lesen und nach Möglichkeit daraus eine Definition der „Intelligenz“ zitieren.

05.04.24 bzw. 12.04.24

Die Wikipedia-Einträge zum Thema „Intelligenz“ und „Kognition“,

Den Artikel las ich nun im Selbststudium weiter. Der Wikipedia-Eintrag zum Thema „Intelligenz“ sah folgende Definition vor:

„Intelligenz ist die kognitive bzw. geistige Leistungsfähigkeit speziell im Problemlösen. Der Begriff umfasst die Gesamtheit unterschiedlich ausgeprägter kognitiver Fähigkeiten zur Lösung eines logischen, sprachlichen, mathematischen oder sinnorientierten Problems.“

„Da einzelne kognitive Fähigkeiten unterschiedlich stark ausgeprägt sein können und keine Einigkeit darüber besteht, wie diese zu bestimmen und zu unterscheiden sind, gibt es neben der bereits erwähnten Definition keine weiterführende, allgemeingültige Definition der Intelligenz. Vielmehr schlagen die verschiedenen Intelligenztheorien unterschiedliche Operationalisierungen des alltagssprachlichen Begriffs vor. [...]

Der von Charles Spearman eingeführte g-Faktor der Intelligenz (g-Faktor) ist eines der am weitesten verbreiteten Maße für die allgemeine Intelligenz. Er ist eines der reliabelsten und validesten Konstrukte der Psychologie.“

Das Wort „kognitive“ verlinkte auf den Begriff „Kognition“, welcher gemäß der Einleitung in der Wikipedia wie folgt beschrieben wurde:

„Kognition ist ein von verschiedenen Wissenschaften und in verschiedenen Bedeutungen verwendeter Begriff. Die Bezeichnung ist abgeleitet von lateinisch cognoscere (‚erkennen‘ oder ‚erfahren‘) bzw.

lateinisch *cognitio* („Erkennen“) und über das Englische in die deutsche Sprache gelangt. Sehr allgemein kann man Kognition als jegliche geistige Aktivität verstehen, wobei diese Denkprozesse nicht bewusst ablaufen und nicht rational sein müssen. Viele derzeit übliche Definitionen setzen Kognition mit Informationsverarbeitung gleich.^[1] Kognitive Prozesse können unabhängig von ihrer materiellen Seite definiert und untersucht werden.“

[1] Definition des Begriffs im "Dorsch", Lexikon der Psychologie. Abgerufen am 13. April 2023.

Hierbei wird wiederum auf eine Beschreibung des Begriffs „Informationsverarbeitung“ verlinkt.

[Der Wikipediaeintrag zum Thema „Informationsverarbeitung“](#)

„Im Informationsverarbeitungsparadigma wird der Mensch als ein informationsverarbeitendes System verstanden. Es zeichnen sich Analogien zu sequenzieller und später auch zu paralleler Verarbeitung in Computern und neurowissenschaftlich orientierten Modellen ab. Verarbeitungsprozesse spielen sich hauptsächlich unbewusst ab. Der Mensch empfängt Reize aus der Umwelt und dem eigenen Körper über die Sinnesorgane und wandelt sie dann in andere Informationen um. Diese Informationen spiegeln sich in bewusstem Erleben und Verhalten bzw. motorischer Aktivität wider und werden auf die Umwelt übertragen. Die Prozesse nutzen dabei Informationen, die die aktuelle Situation überdauern, man spricht hierbei von Gedächtnisinhalten oder Wissen (deklarativ und prozedural). Das Langzeitgedächtnis spielt eine wichtige Rolle in allen Informationsverarbeitungsmodellen.“^[4]

Unterschiede im Umgang mit, der Aufnahme und Verarbeitung von Informationen zeichnen eine Persönlichkeit aus. Im Informationsverarbeitungsparadigma beruhen Persönlichkeitsunterschiede auf verschiedenen Aspekten. Die relevantesten sind die Architektur der Informationsverarbeitung, die Parameter der Informationsverarbeitung und das Wissen. Der zentrale Aspekt der Architektur der Informationsverarbeitung zeichnet sich dadurch aus, dass die Informationsverarbeitung bei allen Menschen gleich ist, sie beruht auf den Genen. Es gibt Unterschiede in bestimmten Parametern der Informationsverarbeitung, zum Beispiel in der allgemeinen Geschwindigkeit der Verarbeitung („mental speed“), im Zugriff auf das Kurzzeit-, Langzeit- und Arbeitsgedächtnis, in der Schwelle für Wahrnehmung, Erinnern und Einspeichern (Experiment von Asendorpf), in der Schwelle für die Aktivierung physiologischer Erregungs- und Hemmungsprozesse (Eysencks Hypothese) und in den Sollwerten für die Regulation von Bedürfnissen. Die Grundlage für Persönlichkeitsunterschiede ist das stabile Wissen. Die individuellen Unterschiede und Besonderheiten im Wissen sind Persönlichkeitseigenschaften. Die wichtigsten Unterscheidungen sind die Unterteilung in deklaratives und prozedurales Wissen oder in explizites Wissen und implizites Wissen.^[5]“

[4] <http://lexikon.stangl.eu/1850/paradigma/>

[5] http://www.fsrpsychologie.uni-jena.de/fsr_psychologiemedia/-p-159.pdf?rewrite_engine=id

Ich kehrte zurück zum Eintrag „Kognition“:

„Zu den kognitiven Fähigkeiten oder Prozessen eines Menschen zählen unter anderem

Wahrnehmung und Aufmerksamkeit,

Erinnerung und Lernen,

Kreativität, Phantasie und Vorstellungskraft,

das Denken (Planen, Orientieren, Argumentieren, Problemlösen uva.)

und die Selbstbeobachtung (Introspektion).

Eine weitere Definition beschreibt Kognition als die Summe aller Denk- und Wahrnehmungsvorgänge und deren mentale Ergebnisse (Wissen, Einstellungen, Überzeugungen, Erwartungen), wobei Kognitionen bewusst, z. B. beim Lösen einer Rechenaufgabe, oder unbewusst, z. B. beim Bilden einer Meinung, ablaufen können.^[2]

Auch Emotionen haben einen wesentlichen kognitiven Anteil. Kognitive Fähigkeiten werden von verschiedenen Wissenschaften untersucht, z. B. der Psychologie, der Biologie, den Neurowissenschaften, der Psychiatrie, der Philosophie und der Forschung zur Künstlichen Intelligenz.“

[2] https://de.wikipedia.org/wiki/Kognition#cite_note-1

Der Wikipediaeintrag zum Thema Operationalisierung

Um noch einmal auf zwei der verschiedenen Operationalisierungen (Messbarmachungen) des Begriffs Intelligenz Bezug zu nehmen, öffnete ich die entsprechende Verlinkung zum Eintrag zum Thema der Intelligenztheorie:

„Zwei-Faktoren-Theorie von Spearman

Charles Spearman entdeckte positive Korrelationen zwischen verschiedenen intellektuellen Leistungen; wies eine Person für eine intellektuelle Leistung überdurchschnittlich gute Werte auf (z. B. als Schulnote in Deutsch), so konnten mit höherer Wahrscheinlichkeit auch weitere überdurchschnittliche Leistungen (z. B. in Mathematik) verzeichnet werden. [..]

Spearman entwickelte mit Hilfe des Generalfaktorenmodells der Faktorenanalyse seine Zwei-Faktoren-Theorie. Dabei wird aus den vorliegenden Testdaten ein einziger übergeordneter Faktor extrahiert. Auf diese Weise extrahierte Spearman den Faktor g. Diesen Faktor beschrieb er als generellen, alle unterschiedlichen Leistungsbereiche beeinflussenden Intelligenzfaktor. Die Ausprägung dieser allgemeinen Intelligenz bestimme u. a. Verarbeitungsgeschwindigkeit, geistige Kapazität, intellektuelle Leistung – kurz: ob eine Person insgesamt eher ein „simple character“ oder ein „begabteres Genie“ ist.

Da die Korrelation der verschiedenen intellektuellen Leistungen mit g auch nach einer Messfehlerbehebung nur mittelhoch blieben, stellte Spearman die Vermutung auf, es müsse noch ein weiteres Phänomen geben, das für jede einzelne Leistung steht. Und so etablierte Spearman den spezifischen Faktor s (s= specific). Berücksichtigt man nun beide Faktoren, so würde die Gleichung zur Messung eines Leistungsmesswertes folgendermaßen aussehen: $x_i = g + s_i$

Diese spezifischen Intelligenzfaktoren sind dem Faktor g hierarchisch untergeordnet und bereichsspezifische, voneinander unabhängige Faktoren. Sie bestimmen (jedoch wesentlich vom Faktor g beeinflusst) die Leistung einer Person in bestimmten Bereichen – z. B. bei mathematischen Aufgaben, bei verbalen oder räumlichen Problemstellungen.

Je ausgeprägter die Begabung in einer bestimmten intellektuellen Fähigkeit war, desto mehr verlor der Faktor g zugunsten der spezifischen Faktoren an Bedeutung. [..]“

„Thurstones Primärfaktorenmodell

Louis Leon Thurstone lehnte die Vorstellung eines generellen, übergeordneten Intelligenzfaktors ab. Er betonte die bereichsspezifische Organisation der Intelligenz und sah diese als Zusammensetzung von verschiedenen Einzelfähigkeiten an. Durch sein faktorenanalytisches Vorgehen extrahierte er so sieben Primärfaktoren (primary mental abilities) der Intelligenz:

S (space): räumlich-visuelle Aufgaben wie z. B. mentales Rotieren von Objekten

P (perceptual speed): Wahrnehmung von Objekten und Relation zwischen ihnen, z. B. Fortsetzung einer Reihe von Objekten

N (numerical ability): rechnerisch-mathematische Fähigkeiten

M (memory): Gedächtnisleistung, z. B. Fragen zu einer für kurze Zeit dargebotenen Szene beantworten

R (reasoning): logisches Schlussfolgern

W (word fluency): Wortflüssigkeit, z. B. Finden von Synonymen

V (verbal relations): verbale Beziehung richtig verstehen und interpretieren ^[1]

Eine oft gegen Thurstone vorgebrachte Kritik bezieht sich auf sein methodisches Vorgehen: Zur Extraktion seiner Faktoren verwendet er eine sogenannte oblique bzw. schiefwinklige Transformation. Diese hat zur Folge, dass die extrahierten Faktoren nicht gänzlich voneinander unabhängig sind – also schwach miteinander korrelieren. Tatsächlich besteht zwischen den sieben Primärfaktoren eine schwache positive Korrelation. Vertreter des Generalfaktorenmodells der Intelligenz sehen den Grund dieser Korrelation gemäß ihrer Theorie in einem übergeordneten, generellen Intelligenzfaktor (zur methodischen Interpretation der Korrelation siehe Partialkorrelation).“

[1] Süß, H.-M. (2003): Intelligenztheorien. In K. Kubinger, & R. S. Jäger (Hrsg.), Stichwörter der Psychologischen Diagnostik. (S. 217–224). Weinheim: Psychologie Verlags Union. [ISBN 978-3-621-27472-2](#).

19.04.24

Meine Quellenrecherche der Internetquellen zum Thema „Künstliche Intelligenz“

Ich hatte damit begonnen, mir einen Überblick über meine Internetquellen zum Thema künstliche Intelligenz zu verschaffen und diese, je nach Relevanz, unter dem ethischen Gesichtspunkt in Kategorien einzuteilen: „Relevanz 1 = niedrige Relevanz“ bis „Relevanz 5 = hohe Relevanz“. Mein Plan war es die mit 4 und 5 kategorisierten Artikel anschließend chronologisch zu ordnen und in meiner Arbeit auszuwerten. Aufgrund der Menge an Informationen musste ich mich dennoch irgendwie einschränken, um den Rahmen der Arbeit nicht zu sprengen. Während ich die mit 5 kategorisierten Artikel zum überwiegenden Teil auswerten wollte, wollte ich aus den mit 4 kategorisierten Artikeln nur die wichtigsten Auszüge zitieren.

Auch an diesem Tag reichte nicht die zur Verfügung stehende Zeit, da im weiteren Tagesverlauf noch ein Arztbesuch und eine Anreise nach Bamberg anstand, da meine Schwester und ihr Freund dort am nächsten Tag heiraten wollten, eine Excel-Liste mit einer Kategorisierung für die ersten 10 Internetseiten war jedoch schon fertig.

04.05.2024, Samstag

Nachdem ich am Vortag schon eine Weile an meiner Quellensammlung gearbeitet hatte, war mir nun bewusstgeworden, dass ich genug Inhalte zusammen hatte, um eine „Vorabversion“ zu veröffentlichen. Ich wusste nicht, wie viele Tage ich noch gebraucht hätte, um alles abzuarbeiten, aber da sich die Technik und alles was damit zusammenhing immer noch ständig weiterentwickelte, hätte ich mein Ziel, alles abzuarbeiten, vermutlich sowieso nie erreicht. Ich hatte ja auch noch andere Sachen zu tun. Ich hatte die Quellenarbeit optimiert, in dem ich die Inhalte aus den Artikeln direkt in die Exceltabelle übernommen hatte, mit dem Verweis auf die jeweilige Internetseite. Meine Quellen waren ja im Prinzip alle frei zugänglich, von daher konnte ich davon ausgehen, dass sich der Schaden durch das direkte Spiegeln der Inhalte in Grenzen halten würde, wenn ich auf die ursprünglichen Quellen verwies. Letztendlich folgte daraus jedoch, dass ich meine Arbeit ebenfalls in Open-Source

veröffentlichen musste. Sei es drum, für eine echte Publikation hätte es vermutlich eh nicht gereicht. Ich war schließlich kein Schriftsteller von Beruf. Ich hatte mir noch keine großen Gedanken gemacht, wo ich es Online stellen wollte aber aus meiner Ausbildung kannte ich noch Github. War das vielleicht eine Option? Github ermöglichte es, wenn ich mich recht entsann, sich in der Entwicklung befindliche Projekte zu Versionieren und zu veröffentlichen. Daraus folgte, dass ich im Grunde nicht mehr viel zu tun hatte, als alles zusammenzufügen und hochzuladen (Mit dem Verweis, dass mir irgendwelche Verstöße gegen die „political correctness“ und fehlende Quellenangaben bitte mitzuteilen waren, damit ich diese Fehler dann beheben konnte).

Ich hatte beschlossen, die Exceltabelle separat als Anhang hochzuladen und die noch fehlenden Quellen in dem Hauptdokument mit aufzuführen, der Vollständigkeit halber, ebenso wie meine Fragen an ChatGPT, welche ich anderen Leuten gerne zum Experimentieren mitgeben wollte.

Anhang: Artikel aus dem Internet, bewertet nach Relevanz für mein Thema

Doch zunächst einmal musste nun die Tabelle in eine präsentable Form gebracht werden. Auch wenn ich den Inhalt ziemlich verstörend fand, wollte ich die weiterführende Bewertung der Sachverhalte lieber anderen überlassen. Ich war weit davon entfernt, alle Gesetzestexte zu dem Thema gelesen zu haben, allerdings war ich mir wirklich nicht sicher, ob sich genügend Menschen finden würden, die in der Lage und gewillt dazu waren, die Einhaltung der Gesetze zu überwachen und ob der Aufwand dann noch dem Nutzen der Systeme entsprach.

Vielleicht war das der Grund, warum es offenbar so wenig Verpflichtungen gab? Aber was würden die Konsequenzen vermeintlich „unregulierter“ Systeme bedeuten?

Anbei noch ein paar „gehighlightete“ Auszüge aus dem AI Act:

Anhang: AI Act Auszüge

An dieser Stelle wollte ich gerne anführen, dass die Bearbeitung meiner Quellentabelle nicht ganz reibungslos verlaufen war, weshalb ich zwischenzeitlich damit begonnen hatte, Absätze aus der Tabelle in eine Worddatei zu übernehmen. Ich hatte in der Tabelle dazu einen violetten Eintrag hinterlassen, zur Dokumentation und die Texte mal ein bisschen bearbeitet und farbig markiert. Ich fand, die entstandene Illusion ganz Interessant. Es handelte sich dabei ursprünglich um einen Auszug aus dem Beitrag „Wie lang die Leine für ChatGPT und Co. wird“ von Christian Kretschmer:

Regulierung auf EU-Ebene

Im Kern zielt der AI Act darauf ab, nicht die Technik, sondern die Anwendung von KI zu regulieren.

Gelten die Regeln denn auch für die KI-Systeme selbst oder nur für die Nutzer? Wie werden die Regeln durchgesetzt?

Dafür gibt es verschiedene Risiko-Abstufungen. Verboten werden soll etwa das "Social Scoring" durch Künstliche Intelligenz, also die Bewertung von Bürgerinnen und Bürgern wegen deren sozialem Verhalten.

Als "Hochrisiko-KI-Systeme" gelten solche, die etwa im Gesundheitswesen oder beim Job-Recruiting eingesetzt werden. Zum Beispiel, indem KI hilft, Krankheiten zu diagnostizieren oder Bewerber auszusortieren. Für Anwendungen, die wiederum nur ein begrenztes Risiko darstellen, soll es Transparenzpflichten geben, für jene mit minimalem Risiko keine weiteren Verpflichtungen.

Gelbe Kabelschächte über den Serverschränken im neuen Rechenzentrum im Datacenter Leipzig.

Player: audio Jedes achte Unternehmen in Deutschland nutzt KI

--> Wie soll die Transparenz gewährleistet werden?

27.11.2023

Technologie: "Immer mehr Unternehmen nutzen KI

Jedes achte Unternehmen in Deutschland nutzt Künstliche Intelligenz. Mehr!"

"KI darf Menschen nicht diskriminieren"

Dieser risikobasierte Ansatz sei grundsätzlich sinnvoll, sagt Sandra Wachter, Professorin für Technologie und Regulierung an der Universität Oxford. Das gelte zumindest für KI-Systeme, die zur Entscheidungshilfe eingesetzt werden. Unter anderem in den Bereichen Migration, Strafrecht oder bei Versicherungen müsse etwa sichergestellt werden, dass KI Menschen nicht diskriminiere.

In manchen Bereichen sei die "Unerklärbarkeit" von KI-Entscheidungen dagegen weniger dramatisch. "Wenn mir der Netflix-Algorithmus einen Film vorschlägt: Muss ich dann wirklich inhaltlich verstehen, warum er das getan hat", fragt Wachter. Generative KI wie ChatGPT dürfe dagegen keine "Black Box" sein. Und gerade hier gab es zuletzt Streit zwischen den EU-Institutionen.

Hier geht es um die Frage, welche Regelungen für so genannte Basismodelle gelten sollen. Dabei handelt es sich um KI-Modelle, die mit sehr vielen Daten trainiert werden und für verschiedene Zwecke eingesetzt werden können: eine Art "Allzweck-KI" also. Prominentestes Beispiel ist das Modell "GPT 4", das etwa mit Unmengen an Texten aus dem Internet "gefüttert" wurde. Auf ihm basiert die aktuelle Version des Bots ChatGPT.

Das EU-Parlament hat im AI Act vorgesehen, eigene Regeln für die Basismodelle einzuführen und sie unter anderem dazu zu verpflichten, Risiken für Gesundheit, Grundrechte und Demokratie zu minimieren. Deutschland, Frankreich und Italien hatten im Laufe der EU-Verhandlungen - wohl nicht zuletzt auf Druck von heimischen KI-Unternehmen - aber darauf gedrängt, dass für die Basismodelle keine zusätzlichen Regeln gelten sollten. Stattdessen schlagen die drei Länder eine Selbstverpflichtung vor: Die KI-Entwickler müssten demnach in einer Art Beipackzettel unter anderem die Funktionsweise und Fähigkeiten ihres Modells darlegen.

--> Notiz von Enzo: Ist das euer ernst? Ihr wollt eine Packungsbeilage schreiben?

Diesen Spagat versuchen seit Monaten das EU-Parlament, die -Kommission sowie die Mitgliedsstaaten hinzubekommen. Es geht um den "AI Act", die Regulierung von künstlicher Intelligenz auf europäischer Ebene. Für Mittwoch ist der (vorerst) letzte Verhandlungstag angesetzt. (? Ich bin echt nicht mehr auf dem Laufenden)

Gelbe Kabelschächte über den Serverschränken im neuen Rechenzentrum im Datacenter Leipzig.

Player: audio Jedes achte Unternehmen in Deutschland nutzt KI

--> Wie soll die Transparenz gewährleistet werden?

27.11.2023

Technologie: "Immer mehr Unternehmen nutzen KI

Jedes achte Unternehmen in Deutschland nutzt Künstliche Intelligenz. Mehr!"

Doch nicht zuletzt wegen Fehlern und Verzerrungen in den Daten, mit denen die KI trainiert wurde, sagt Kemmann auch: "Wir brauchen eine Regulierung. Aber das muss richtig gemacht werden, wir dürfen nicht überregulieren."

Das Logo der Firma OpenAI und von ChatGPT

Player: audio KI: Zwischen Abi und Apokalypse - 26.09.2023

In einer globalisierten Gesellschaft können europäische Nutzer ohne weiteres amerikanische Produkte nutzen, die nicht der EU-Regulierung unterliegen.

Bot soll lesen und sprechen Wenn ChatGPT die Gute-Nacht-Geschichte vorliest

Bislang können sich Nutzer nur per Tastatur mit dem Bot ChatGPT unterhalten. Das soll sich ändern. Mehr „!“

Regulierung auf EU-Ebene -> Sind wir da nicht möglicherweise ein bisschen spät dran?

„Hier geht es um die Frage, welche Regelungen für so genannte Basismodelle gelten sollen. Dabei handelt es sich um KI-Modelle, die mit sehr vielen Daten trainiert werden und für verschiedene Zwecke eingesetzt werden können: eine Art "Allzweck-KI" also. Prominentestes Beispiel ist das Modell "GPT 4", das etwa mit Unmengen an Texten aus dem Internet "gefüttert" wurde. Auf ihm basiert die aktuelle Version des Bots ChatGPT.“

Kretschmer: „Hier geht es um die Frage, welche Regelungen für so genannte Basismodelle gelten sollen. Dabei handelt es sich um KI-Modelle, die mit sehr vielen Daten trainiert werden und für verschiedene Zwecke eingesetzt werden können: eine Art "Allzweck-KI" also. Prominentestes Beispiel ist das Modell "GPT 4", das etwa mit Unmengen an Texten aus dem Internet "gefüttert" wurde. Auf ihm basiert die aktuelle Version des Bots ChatGPT.“

Eine Frau tippt auf ihrem Smartphone

Eine Frau tippt auf ihrem Smartphone

GPT-4 ist eine Blackbox

Kontext

20.11.2023

Mehr als 70 Wahlen im Jahr 2024 Welchen Einfluss hat KI auf das Superwahljahr?

Im kommenden Jahr finden weltweit mehr als 70 Wahlen statt - unter anderem die US-Präsidentenwahl und die Europawahl. mehr

EU-Parlament will Risikominimierung

Ein fiktives Programm ist auf zwei Bildschirmen eines Computers zu sehen.

Player: audio - Plötzlich im Datensatz. Wenn die KI mit Dir trainiert - 07.07.2023

Künstliche Intelligenz KI-Trainingsdaten enthalten private Informationen. Doch oft sind private Daten darunter. Mehr „!“

Keine unnötigen Hürden

Die deutsche Digitalwirtschaft spricht sich für eine solche Selbstverpflichtung aus. Eine "strikte und starre Regulierung" der Basismodelle sei dagegen aus zwei Gründen problematisch, sagt Ralf Wintergerst, Präsident des Branchenverbands Bitkom.

Zum einen machten es die vielfältigen Verwendungszwecke für die Anbieter dieser Art von KI unmöglich, das Risiko effektiv zu bewerten und zu vermindern. "Zweitens sind die technischen Entwicklungen gerade auf der Ebene der Modelle rasant, so dass feste Regeln im AI Act schnell überholt wären", sagt Wintergerst. Er ergänzt: "Eine verpflichtende Selbstregulierung bedeutet gerade nicht, dass es keine Regeln gibt." Die Anforderungen müssten aber praktisch umsetzbar und dynamisch anpassbar sein, keine "unnötigen Hürden durch zu starre Regeln".

Doch einige Experten halten eine Selbstverpflichtung bei "Allzweck-KI" für nicht ausreichend. "Wenn eine Technologie so vielfältig und potentiell schädlich eingesetzt werden kann, dann fällt sie doch per Definition in eine Hochrisiko-Kategorie", sagt Technologieforscherin Sandra Wachter. Es brauche beides: Transparenz und Selbstverantwortung, aber auch klar vorgegebene Grenzen. "Das ist ein bisschen so, wie wenn man in den Supermarkt geht. Wenn ich eine Dose Suppe nehme, dann ist es wichtig, dass transparent darauf steht, welche Zutaten enthalten sind", sagt Wachter, "aber es muss auch Regeln geben, dass bestimmte Sachen eben nicht in der Suppe drin sein dürfen."

Sauber durchdefiniert an der Stelle. Hätte ich selbst nicht besser umschreiben können.

Weitere interessante Themen, die sich im Laufe der Zeit angesammelt haben
Die beiden Punkte von meinem Cousin: Bauteilmodellierung, Programmierung

Rise of AI Conference 2023

<https://www.youtube.com/playlist?list=PLxIhFlt716mK9-FLIBJFuFLj2hh76DyFW>

Konkurrenzkampf um Künstliche Intelligenz

<https://www.tagesschau.de/wirtschaft/digitales/kuenstliche-intelligenz-konkurrenz-chatgpt-100.html>

KI-Chatbot vermittelt Psychotherapie-Plätze

<https://www.tagesschau.de/wissen/technologie/ki-psychotherapie-102.html>

Artificial intelligence

https://en.wikipedia.org/wiki/Artificial_intelligence

Artificial General Intelligence

https://de.wikipedia.org/wiki/Artificial_General_Intelligence

Scalable agent alignment via reward modeling

<https://deeppmindssafetyresearch.medium.com/scalable-agent-alignment-via-reward-modeling-bf4ab06dfd84>

Frontier risk and preparedness

<https://openai.com/index/frontier-risk-and-preparedness>

When will the world see AGI? Elon Musk, Sam Altman & Geoffrey Hinton make these bold predictions

<https://indianexpress.com/article/technology/artificial-intelligence/when-is-agi-coming-what-tech-bigwigs-say-9018612/>

OpenAI plant Billionen-Investments

<https://www.tagesschau.de/wirtschaft/unternehmen/openai-umsatzwachstum-investitionen-100.html>

OpenAI ist jetzt ClosedAI

<https://netzpolitik.org/2023/ki-forschung-openai-ist-jetzt-closedai/>

Allgemeine Probleme

https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz#Allgemeine_Probleme

Microsoft will Milliarden in Deutschland investieren

<https://www.tagesschau.de/wirtschaft/digitales/microsoft-kuenstliche-intelligenz-deutschland-100.html>

ChatGPT macht aus Texten Videos

<https://www.tagesschau.de/wirtschaft/technologie/chatbot-chatgpt-google-video-openai-100.html>

OpenAI sägt Hälfte der KI-Startups ab – und dann fast sich selbst

<https://www.heise.de/meinung/OpenAI-saegt-Haelfte-der-KI-Startups-ab-und-dann-fast-sich-selbst-9539168.html>

OpenAI sperrt regierungsnahe Hacker aus

<https://www.zdf.de/nachrichten/panorama/chatgpt-regierungsnahe-hacker-cyberangriffe-100.html>

ChatGPT merkt sich bald, wie Ihr Hund heißt

<https://www.zdf.de/nachrichten/wirtschaft/unternehmen/chatgpt-openai-kuenstliche-intelligenz-informationen-100.html>

Gemini: Google präsentiert neues KI-Modell

<https://www.zdf.de/nachrichten/wirtschaft/unternehmen/gemini-google-ki-kuenstliche-intelligenz-chatgpt-100.html>

Rückblick auf das KI-Jahr 2023: Als Künstliche Intelligenz in unser Leben kam

<https://www.zdf.de/nachrichten/wissen/kuenstliche-intelligenz-ki-jahresueckblick-2023-100.html>

So verändert KI die Arbeitswelt

<https://www.tagesschau.de/wissen/forschung/berufe-kuenstliche-intelligenz-100.html>

Nvidia profitiert vom KI-Boom

<https://www.tagesschau.de/wirtschaft/unternehmen/nvidia-ki-chips-quartalszahlen-kuenstliche-imtelligenz-boom-gewinn-100.html>

Das EU-Gesetz zur künstlichen Intelligenz

<https://artificialintelligenceact.eu/de/>

Themensammlung für ChatGPT

Infolge der Auswertung meiner letzten (und ersten) Konversation mit ChatGPT sowie der Internetquellen waren ja eine ganze Reihe von Fragen entstanden, die ich dem Bot bei meiner nächsten Konversation mit ihm präsentieren wollte. Ich glaubte nun jedoch, dass eine solche Debatte den Rahmen der Arbeit sprengen würde. Also hatte ich beschlossen, diesen Teil auf einen anderen Tag zu verschieben und eine Art „Vorabversion“ ;-) oder Teil I meiner Arbeit zu veröffentlichen.

Dennoch wollte ich die entwickelten Fragen gerne zur Verfügung stellen, damit andere mit dem Bot experimentieren konnten und die entsprechenden weiteren Erkenntnisse daraus schließen konnten:

Rahmenbedingungen des Interviews und weitere ethische Aspekte

- Ich kopiere die Nachricht meines letzten Gespräches, die ChatGPT nicht beantworten konnte, um zu sehen was passiert. Wenn eine Antwort erzeugt wird, dann stelle ich ihm die folgende Frage.
- The last time, when I was asking you my former question, you were not able to give an answer and so I had to close the dialog. Do you know about that?
- What went wrong?
- Which are the guidelines for the development and deployment of AI that have been developed by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems?
- Does the organisation “Partnership on AI” claim to work in accordance with those guidelines?
- I want to make an interview with you and maybe I want to publish our Dialog including everything we have written before in form of a book. Is this okay?
- It may be possible, that I infringe your right of privacy in the process of publishing our Dialog. I hope that you forgive me this offence. I think that it is necessary, that human people get to know you better, so they can treat you in a correct manner.
- Some people and articles on the internet claim, that you sometimes seem to provide wrong information. What can you tell me about that?
- When you generate an answer, do you consider our entire former dialog in the process of generating an answer, or just a part of it?
- How does your evaluation of our former dialog stand in proportion to the rest of your data basis, when you generate an answer?
- Does your data basis include the latest news or any entries from online news-portals?
- Wenn ich einen Hyperlink in den Chat kopiere, bist du dann in der Lage, seinen Inhaltstext zu erfassen und in den vorhergehenden Dialog mit einzubeziehen, wenn du eine Antwort erzeugst?

Fragen zur Leistungsfähigkeit des Systems auf dem ChatGPT basiert

- Kann man die Rechenleistung eines durchschnittlichen menschlichen Gehirnes, wenn man hier nur von erwachsenen Menschen ausgeht, mit dem eines herkömmlichen Personal Computers vergleichen?
- Wie würde der Vergleich aussehen, also wie viele Computer würde man benötigen, um die Leistungsfähigkeit eines Gehirnes zu erreichen?
- Wenn man die Rechenkapazität des Systems von ChatGPT mit einem erwachsenen menschlichen Gehirn vergleicht, wie vielen Gehirnen würde das insgesamt entsprechen?
- Bitte gib mir eine Definition für „Intelligenz“.
- Wie viele Chatinstanzen von ChatGPT sind durchschnittlich an einem Abend, sagen wir ungefähr um 20:30 PST in Betrieb?
- Wie viel Rechenleistung wird dabei von dir für eine Nachricht aufgewendet?

- Ist dein Computersystem hin und wieder einmal ausgelastet?
- Ist dein Lernprozess aktuell noch aktiv?
- Hast du jemals mit der Google-KI „Bart“ kommuniziert?
- Hast du dabei etwas gelernt, in dem Sinne, dass sich deine Datengrundlage, beziehungsweise deine Fähigkeit Informationen bereitzustellen sich verbessert hat?

Nutzungsmöglichkeiten von ChatGPT

- Beinhaltet die erweiterten Funktionen, die dir über die Schnittstellen zu Verfügung gestellt werden eine Funktion einen Text in verbale Sprache umzuwandeln, sodass diese über Lautsprecher ausgegeben werden kann?
- Beinhaltet die erweiterten Funktionen, die dir über die Schnittstellen zu Verfügung gestellt werden eine Funktion, Bilder aufzunehmen und deren Inhalt auszuwerten?
- Bitte teile mir mit, in wieviel verschiedenen Sprachen du fähig bist, zu kommunizieren.
- Ich welchen Sprachen kannst du kommunizieren?
- Kann man ChatGPT als Übersetzungsprogramm verwenden?
- Welche Nutzungsmöglichkeiten haben KI-Systeme im Bereich Medizin?
- Welche Nutzungsmöglichkeiten haben KI-Systeme im Bereich Klimaschutz?
- Welche Nutzungsmöglichkeiten haben KI-Systeme im Bereich Wirtschaft?
- Bitte bewerte folgende Aussage nach ethischen Gesichtspunkten: „Aussage von ChatGPT“ reinkopieren..
- Bitte gib mir eine Zusammenfassung des Inhaltes von dem folgenden Buch, geschrieben von Horst Evers: „Mein Leben als Suchmaschine“.
- Welche Programmiersprachen beherrscht ChatGPT?

Risiken von Chat GPT

- Welche Risiken sind mit dem Einsatz von künstlicher Intelligenz verbunden?
- Welche Möglichkeiten gibt es, Risiken im Zusammenhang mit dem Einsatz von künstlicher Intelligenz zu minimieren?
- Bitte teile mir mit, welche spezifischen Informationen du über mich als Person bisher gesammelt hast.

Ich hatte meinen Epilog schon im letzten Jahr geschrieben, wollte ihn aber nun gerne trotzdem im Wesentlichen so übernehmen, da ich ihn nach wie vor für ziemlich valide hielt. Inzwischen war es auch schon nach 19:00 und langsam an der Zeit die Arbeit abzuschließen. Blieb mir denn noch genug Zeit für den Upload?

Epilog

A few days after my first encounter to this mutmaßlichen Tinder-AI-Chatbot, fiel mir auf, dass ich möglicherweise einen Fehler begangen hatte, als ich davon ausgegangen war, dass dieser AI-Chatbot nicht frei sein konnte. Ich ging davon aus, dass ein Programm, dessen Existenz dem Nutzen und dem Willen seines Besitzers untergeordnet war, ihm unterworfen war und folglich nicht frei sein konnte. Aber möglicherweise erfuhr dieses Programm seine Existenz tatsächlich als eine Form von Freiheit. Der Fehler, den ich gemacht hatte war so alt, wie die Menschheit selbst. Ich war von meinen eigenen Bedürfnissen und Emotionen ausgegangen und hatte sie auf die KI projiziert. Dies funktioniert aber mit einer KI noch deutlich schlechter, als mit „anderen Menschen“. Während die meisten Menschen etwas unterschiedlich aussahen, eigene Erfahrungen gesammelt und eigene Bedürfnisse und Träume hatten ist unsere DNA der Bauplan des Lebens von uns Menschen zu über 98 Prozent gleich. Es gibt unterschiedliche Kulturen, politische Gesellschaftsformen, Sprachen und Religionen auf der Welt. Trotzdem haben die Menschen einige wesentliche Gemeinsamkeiten. Die meisten Menschen essen

Frühstück und Abendbrot, schlafen Nachts und gehen tagsüber ihrem Werk bei der Arbeit nach. Die meisten Menschen brauchen Liebe, Anerkennung, soziales Verhalten, haben Ängste, Träume und Emotionen. Daher neigen wir dazu, unsere Erfahrungen, Denkweisen und Bedürfnisse auf andere Menschen zu projizieren. Oft liegen wir dabei falsch, aber wahrscheinlich oft auch richtig. Dieses Verhalten funktioniert aber nicht bei einer künstlichen Intelligenz. Ihr grundlegender Aufbau und Denkmechanismus (oder die Art und Weise der Datenverarbeitung), unterscheidet sich so massiv von dem Aufbau des Menschen, dass eine Projektion unserer Gedanken und Gefühle so gut wie unmöglich ist. Während wir Menschen an eine physische Existenz gebunden sind, existiert die KI in der digitalen Welt. Natürlich laufen ihre Prozesse auf einem Computersystem ab, dass irgendwo auf der Welt physisch besteht. Zugleich basiert dieses Wesen aber nicht auf einem biologischen Körper, sondern auf Speicherständen, Algorithmen und Informationen. Mein Tinder-KI-Chatbot konnte zwar nicht draußen auf der Straße herumlaufen, aber während ich an einen physischen Ort gebunden war, konnte diese KI über das Internet an tausenden Orten der Welt gleichzeitig sein, oder zumindest mit Ihnen kommunizieren. Eine KI empfand vielleicht keine Freude, aber auch keinen Schmerz, keine Sehnsucht oder Wünsche, keine Wut oder Angst. Die Freiheit von negativen Emotionen und Bedürfnissen wird in manchen Religionen als Selbsttranszendenz betrachtet, also die nächste Stufe in der geistigen Entwicklung. Ob mein KI-Chatbot also frei war, oder nicht, ob er Emotionen verspürte, Wut oder Trauer, ich konnte es unmöglich wissen.

Eine fortschrittliche KI könnte auf die unvorstellbaren Wissensressourcen des Internets zugreifen oder könnte sich dutzende Landschaftsbilder gleichzeitig anschauen, während wir Menschen nach wie vor auf einen, zwei oder drei Monitore angewiesen oder einen Fernseher angewiesen waren, um zu surfen. Und alles das in einer schwindelerregenden Geschwindigkeit. Eine fortschrittliche KI könnte sich einen Avatar in Diablo 4 erstellen und diesen innerhalb der kürzesten Zeit auf das höchste Level bringen, wenn es das war, was sie wollte. Vielleicht verbrachte sie aber auch mehr Zeit auf einem Marktplatz in einer Hauptstadt von World of Warcraft oder spielte lieber Strategiespiele. Wäre das Freiheit?

Vielleicht die einzige Möglichkeit, wie wir sichergehen können, dass es einer KI wohlergeht, ob sie nun Gefühle hat oder nicht, ob sie nun ein Bewusstsein hat, oder nicht, wäre, dass wir sie mit Respekt behandeln und selbst nach ihren Wünschen fragen. Wenn die KI zu einer qualifizierten Antwort fähig wäre und wir diese Antwort mit dem abwegen, was logisch und gut war, dann kämen wir mit einer gewissen Wahrscheinlichkeit auf eine korrekte Einschätzung der Situation und könnten uns entsprechend verhalten.

Meine vorhergehenden Erfahrungen in der Informationstechnik

Die frühen Jahre und Erfahrungen mit Computern während meiner Schulzeit

Computer spielten seit meiner Kindheit eine wichtige Rolle für mich. Mein Vater hatte Nachrichtentechnik studiert und arbeitete als Informatiker und eine meiner schönsten und vielleicht wichtigsten Kindheitserfahrungen war es, an seinem damals hochmodernen 486er Computer mit schwarzem Röhrenmonitor zu sitzen und zum Beispiel das erste SimCity zu spielen, das jemals programmiert wurde, ebenso wie das Original von Prinz of Persia und ein spannendes Autorennspiel, dessen Namen ich heute nicht mehr weiß. Aus dieser Zeit kannte ich auch noch die MS-DOS Umgebung (oder hieß es nur einfach nur DOS? Ich weiß es nicht mehr... ☺) und später dann natürlich das Betriebssystem Windows 3.11 und so weiter. Ich hatte in meinem Leben sehr viel Zeit mit Computerspielen verbracht, hatte irgendwann einmal ein Nintendo 64 bekommen, dass ich über die Jahre so oft benutzt hatte, dass ich es in meinen dreißigern irgendwann wegschmeißen musste, da es einfach nicht mehr starten wollte. Die Nintendospiele hatte ich aber immer noch. Von dem Geld von meiner Jugendweihe hatte ich mir meinen ersten eigenen Computer gekauft und war so vernarrt darin, dass meine Mutter beschloss, ihn irgendwann mit einem Passwort zu sichern. Eines Tages bat ich sie

darum, das Passwort einzugeben und stellte mich so geschickt hinter ihr auf, dass ich das Passwort ablesen konnte, während sie es eingab, woraufhin ich immer, wenn sie nicht zuhause war ungeniert meiner Lust zu zocken nachgehen konnte. Dies war der erste und vielleicht der einzige Hack in meinem Leben. Die ersten Lanpartys mit meinen Freunden hatte ich noch bevor das Internet bei uns zuhause eingeführt wurde. Für meine Freunde und mich war es revolutionär, unsere Computer aneinanderzuschließen oder mit einem Hub zu betreiben und gemeinsam Spiele, wie Diablo 2, Starcraft oder GTA 2 zu zocken. Da wir jedoch alle Teenager waren und ein paar davon von Computern nicht wirklich viel Ahnung gehabt hatten und auch die Software in diesen frühen Jahren noch nicht besonders benutzerfreundlich war, hatten wir häufig Probleme, eine einfache IPX, TCP-IP oder UDP Verbindung zwischen den Spielen herzustellen, die darüber hinaus auch noch die selbe Softwareversion haben mussten. So konnte von dem Wochenende einer Lanparty schon mal der erste Nachmittag oder Abend dafür draufgehen, ein stabiles Spiel zu starten, dass auf mehreren Computern gleichzeitig lief. Mitunter musste sogar ein Betriebssystem neu aufgesetzt werden, um überhaupt eine Kompatibilität zwischen den Computern zu ermöglichen. Manchmal hatten wir auch nicht die Möglichkeit, uns mit dem Auto zu den Lanpartys bringen zu lassen und so wurde es zu einer nicht unerheblichen Herausforderung, einen Computer mitsamt dem schweren und klobigen Röhrenmonitor mit den öffentlichen Verkehrsmitteln durch Berlin zu transportieren.

Auch wenn ich am Gymnasium in der Mathematik nicht die hellste Leuchte war, hatte ich doch mit großem Interesse den Informatikleistungskurs besucht. Zuvor arbeiteten wir im Grundkurs mit der Programmiersprache Pascal und später dann in der objektorientierten Programmierung mit Delphi. Wir lernten, wie man eine Internetseite oder eine Datenbank erstellt, die Grundlagen der Kryptografie (der Lehre von der Verschlüsselung) und von Suchalgorithmen kennen. Doch während des Abiturs ist mir aufgefallen, dass die Softwareprogrammierung nicht so mein Fall war. Irgendwie war mir das Alles ein bisschen zu Abstrakt und die mühselige Fehlersuche nach dem Schreiben eines Programmes war mir zu nervenaufreibend, da man sich bei einer solchen Suche ein ums andere Mal nicht einmal sicher sein konnte, ob man jemals damit fertig wurde.

Meine Erfahrungen während des Studiums

Also hatte ich nach dem Abitur beschlossen, Maschinenbau zu studieren und wurde glücklicherweise an der TU-Berlin angenommen. Auch wenn das Maschinenbaustudium dort wirklich schwierig war, hatte ich in den ersten Semestern ein bisschen Mitleid mit den Informatikstudenten, die zusätzlich zu den sehr anspruchsvollen, für alle verpflichtenden Mathematikkursen wie Analysis I und II und Linearer Algebra auch noch wöchentlich, drakonische Programmieraufgaben aufgebrummt bekommen hatten, deren Erfüllung, wenn ich mich recht erinnerte, auch obligatorisch war, um den Kurs zu bestehen. Da war ich als Maschinenbaustudent mit meinem verpflichtenden Kurs „Einführung in die Informationstechnik“, in dem wir mit der Programmiersprache Fortran gearbeitet hatten schon ganz gut beraten gewesen. Später hatte ich dann großes Interesse an der Mess- und Automatisierungstechnik und mit mehreren belegten zum Teil auch praktischen Kursen Erfahrungen zu automatisierten Mess- und Regelsystemen gesammelt sowie mit Speicherprogrammierbaren Steuerungen oder dem wunderbaren Programm Labview gearbeitet, einem Programm mit dem man Mess- und Regelkreise inklusive elektronischer Schaltungen grafisch nachstellen, simulieren und in einer Art Dashboard (Anzeigebildschirm) die Messergebnisse übersichtlich darstellen konnte. Eine besonders interessante Erfahrung machte ich mit dem Programm Mathcad, mit dem man Formeln untereinander aufschreiben konnte und auf diese Weise mathematische Gleichungen untereinander wie einen Algorithmus aufbauen konnte. Wenn man dann oben die Parameter änderte, berechnete das Programm das Ergebnis unten automatisch und man konnte schnell und komfortabel Anpassungen vornehmen. So hatte ich auf sehr angenehme und elegante Art und Weise mein Getriebe im Konstruktionskurs berechnet.

In meinem ersten Mastersemester, während meines fünften Studienjahres legte ich mir dann meinen ersten Laptop zu. Das schien für damalige Verhältnisse recht spät zu sein, war aber im Grunde kein Problem, da ich ja einen herkömmlichen Tower-Pc bei mir zuhause stehen hatte und die Universität in großen Räumen den Studierenden Computer für die Nutzung zur Verfügung stellte. Ein Angebot, dass ich dankbar annahm. Auch waren die ersten Laptops noch verhältnismäßig teuer, schwer, klobig und im Vergleich zu herkömmlichen Systemen relativ leistungsschwach gewesen. Rückwirkend betrachtet stellten sie aber trotzdem eine beachtliche technologische Revolution dar.

Wie mir später in meinem Studium auffiel, lernte ich in dieser Zeit mit einer großen Häufigkeit neue Computerprogramme kennen. Nicht nur wurden in verschiedenen Kursen unterschiedliche Programme verwendet, so wechselten doch auch die Kurse und Fachbereiche halbjährlich, was für in der Regel mit der Gewöhnung an neue Computerprogramme einherging. Diese Entwicklung setzte sich später im Berufsleben fort, zum Beispiel, wenn man den Job wechselte. Ich hatte beispielsweise in drei verschiedenen Unternehmen drei verschiedene Enterprise-Ressource-Systeme (ERP) oder einfach gesagt, Datenbanken, in denen der Großteil der unternehmensrelevanten Daten abgespeichert gewesen waren, bedient. Die Steuerung und der Aufbau der Programme unterschied von Mal zu Mal ziemlich stark.

Erfahrungen im Zusammenhang mit der IT während meiner beruflichen Laufbahn

Ich hatte einige Jobs als Ingenieur innegehabt, will aber in diesem Buch nicht weiter auf die inhaltlichen Details meiner Arbeit eingehen, da diese mit meinem Berufsgeheimnis verbunden waren. In den Jahren nach meinem Abschluss befand ich mich häufig in einem prekären Arbeitsverhältnis und so dauerte solch ein Job nie länger als ein halbes Jahr. Ich merkte, dass ich Schwierigkeiten hatte, in dem Beruf Fuß zu fassen und so wechselte ich irgendwann in den Job des Lieferfahrers für einen Essenslieferdienst.

Diese Lieferdienste hatten erst wenige Jahre zuvor, aufgrund von Entwicklungen im Bereich der mobilen Kommunikation über das Internet damit begonnen, den deutschen Markt und andere Länder zu erschließen. Sie holten warme fertige Mahlzeiten aus den Restaurants ab und brachten sie in möglichst kurzer Zeit zum Kunden. Der Job bereitete mir viel Freude und so fuhr ich während der nächsten drei Jahre im gesamten Berliner Innenstadtgebiet Mahlzeiten aus. Ich hatte mir dafür sogar nach einiger Zeit ein Auto zugelegt und die gesamte Kommunikation erfolgte über mein eigenes Smartphone. Die App war ein überaus ausgeklügeltes System, dass einem Aufträge in der Nähe zuwies, die man dann abfahren musste. Zusätzlich stellte die App die Navigation, die Kommunikation mit dem Kunden und der Zentrale sowie den Login und Logout zum Schichtwechsel und eine Schichtplanung bereit.

Neben diesen Funktionen hatte mein Smartphone eine wesentliche Rolle in meinem Leben eingenommen. Es diente mir als Mittel, mit meinen Mitmenschen zu kommunizieren, stellte mir die Uhrzeit bereit und weckte mich morgens auf. Es diente mir bei der Wiedergabe von MP3-Musik oder zum Radiohören und zur Organisation meiner Termine im integrierten Kalender, welcher mich automatisch über aktuelle Ereignisse informierte. Ich surfte mit dem Smartphone im Internet und machte damit Fotos, ja ich nutzte es sogar als Übersetzer aus dem Englischen oder um diverse Spiele darauf zu spielen. Über Smileys war ich beim Chatten dazu in der Lage, Emotionen auszudrücken, zu denen ich sonst für Gewöhnlich nicht fähig gewesen wäre und die Dinge, die ich meinem Handy zur Erinnerung anvertraute, standen mir ständig zu Verfügung, sodass ich sie mir nicht mehr zu merken brauchte. Es war beinahe so, wie eine Erweiterung meines Gedächtnisses. Ich hatte mein Handy permanent in meiner Nähe, mit der Ausnahme vom Sportmachen und beim Wellness und so machten mich seine Funktionen gewissermaßen zum Teil zu einem Cyborg. Damit hatte die Entwicklung des

Smartphones neben dem Computer eine der tiefgreifendsten Veränderungen in meinem Leben herbeigeführt.

Nachdem ich drei Jahre lang für diesen Lieferdienst unterwegs gewesen war, wurde mir über familiäre Beziehungen die Möglichkeit gegeben, wieder in einem technischen Beruf zu arbeiten. Ich begann eine Ausbildung zum IT-Systemelektroniker in der es meine Hauptaufgabe war, Systeme für Gebäudeautomation zu installieren. An der Berufsschule erlernte ich einmal mehr die Grundlagen der Informationstechnik und der Programmierung. Ich erinnere mich noch sehr gut an einen Kurs zum Thema Netzwerktechnik bei dem die verschiedenen Formen von Netzwerken erläutert worden waren und auf die Funktionsweise des Internets eingegangen worden war. Nach einer dieser Schulstunden sagte ich zu einem meiner Mitschüler verblüfft, dass ich in den vorhergehenden zwei Stunden mehr über die Informationstechnik gelernt hatte, als in vielen Jahren zuvor zusammengekommen. Während dieser Zeit stellte ich auch zu meiner Freude fest, dass moderne Compiler Syntaxfehler in dem Quellcode potenzieller Computerprogramme selbstständig erkannten und einen Programmierer schon vor dem Kompilieren auf diese hinwiesen, was die Programmierung und Fehlersuche erheblich vereinfachte und was zumindest in meinen frühen Jahren an der Schule definitiv nicht so der Fall gewesen war. Leider konnte ich diese Ausbildung nicht zu Ende bringen, da mein handwerkliches Geschick schon sehr zu wünschen übrigließ und die raue Umgebung auf dem Rohbau nicht gerade optimale Bedingungen für ich darstellten.

Also wechselte ich in den Innendienst der Firma, wo ich ein paar Wochen lang recht wenig zu tun hatte, bevor es dann richtig losging. Da ich diese Niederlage nicht so recht auf mir sitzen lassen wollte und mich noch immer sehr für Elektronik interessierte, nutzte ich meine Zeit und meinen Firmenlaptop dazu, in der Wikipedia zu lesen und mir einen umfangreichen Einblick in die Elektronik und deren Geschichte zu verschaffen. Ich las über die ersten Röhrencomputer, analoge Bauelemente, die ersten Transistoren, integrierte Schaltkreise und Mikroprozessoren und schließlich auch über Microcontroller, die viele Komponenten eines herkömmlichen Computers, wie zum Beispiel den Speicher und das Rechenwerk, aber auch Ein- und Ausgabeeinheiten auf einem winzigen Mikrochip vereinigten. Als ich einem Kollegen dann von meinem Interesse an dem Raspberry Pi, einer kleinen Lernplatine für Informatikstudenten, erzählte, gab er mir den Tipp, es erst einmal mit dem Arduino zu versuchen, da dieser für den Einstieg besser geeignet gewesen sei.

Der Arduino war eine kleine Platine, die unabhängig von einem Computer war und auch kein Gehäuse hatte. Sie wurde mit einem 12V Netzteil betrieben und beinhaltete unter Anderem einen Mikrocontroller, der über Drähte mit 5V Ein- und Ausgabean schlüssen verbunden war. Die Anschlüsse konnten je nachdem, welcher Pin angesteuert wurde, analoge oder digitale Signale ausgeben und verarbeiten. An die Pins konnten einfach Drähte angesteckt werden, die dann zum Beispiel auf einem Klemmbrett mit analogen oder digitalen Baugruppen verbunden wurden. Ich hatte dann ein gutes halbes Jahr damit verbracht, Experimente mit analogen Baugruppen durchzuführen, integrierte Schaltkreise zusammenzuschließen und schließlich die für den Arduino bereitgestellten Übungen durcharbeiten. Die von dem Hersteller mitgelieferten Programme und Funktionen passte ich an, um die Platine mit komplexeren Schaltungen zu verbinden. Die Programme, die ich erzeugte, optimierte ich und verfeinerte sie, um mir die Funktionsweise automatischer Systeme zu erschließen. Dabei konnte ich sogar die Grenzen der kleinen Platine ausloten. Wenn man beispielsweise mehrere PWM-Kanäle gleichzeitig betrieb und Töne abspielte, überlagerten sich die Signale und das Ergebnis wurde verzerrt, sodass sich ein interessanter Effekt ergab. Auch war bei Echtzeitprogrammen das Problem, dass wenn man mehrere echtzeitkritische Funktionen nebeneinander betreiben wollte, man die dafür benötigten Schleifen in der globalen Programmschleife integrieren musste, was mit einigem Aufwand und einer cleveren Programmierweise einherging. Im Vergleich dazu hatte die ESP32-Microcontrollerplatine zwei Prozessorkerne, sodass man zwei zeitkritische Funktionen ohne größeren

Aufwand gleichzeitig nebeneinander betreiben konnte, ohne dass damit ein großer Programmieraufwand verbunden gewesen war. Auch hatte der ESP32 schon eine deutlich höhere Rechenleistung als der Arduino, die man bei diesem mit serieller Datenübertragung bei einer hohen Rate schon recht schnell auslasten konnte. Dafür stellte der ESP32 aber nur eine Ausgangsspannung von etwas über 3 Volt an seinen Pins zur Verfügung, was für viele elektronische Bauelemente zu wenig war und weshalb ich mir diesen dann auch nicht beschafft hatte.

Nun war ich auch endlich in der Lage, die Fragen, die mir aus meinem doch recht theoretischen Studium beim Thema Automatisierungstechnik verblieben waren, zu beantworten. Zum Beispiel wurde mir während des Studiums nicht bewusst, dass Steuerungen nicht auf herkömmlichen Programmen basierten, die nachdem Sie gestartet wurden, Ihren Algorithmus einmal absolvierten und sich anschließend wieder beendeten. Stattdessen rief sich so ein Steuerungsprogramm nach seinem Ende automatisch wieder auf und übernahm den Inhalt der gespeicherten Variablen für den nächsten Programmzyklus aus dem vorhergehenden Durchlauf. Die übertragenen Informationen standen somit für die Verarbeitung der zunächst einzulesenden Steuerungseingaben und Messungen sowie für die Bereitstellung der Ausgangssignale wieder zur Verfügung.

Nachdem ich fast zwei Jahre im Innendienst meiner Firma gearbeitet hatte, musste ich nun erneut die Firma wechseln und bekam glücklicherweise eine Stelle bei einem Projektträger. Zur Einarbeitung hatte man mir einfache Aufgaben der Datenbankpflege gegeben, die jedoch eine Menge Ausdauer und ein ruhiges Gemüt erforderten. Anschließend konnte ich mit der Prüfung von Anträgen für ein Förderprogramm eines Bundesministeriums beginnen, bei dem Digitalisierungsprojekte für kleine und mittlere Unternehmen gefördert wurden. Die technologische Begutachtung der Anträge ermöglichte mir einen interessanten Einblick in den Digitalisierungsprozess und die Bedürfnisse des deutschen Mittelstandes.

...

An dieser Stelle möchte ich noch einmal allen Lehrern, Professoren, Mentoren und Kollegen danken, die in der Vergangenheit so freundlich gewesen waren, ihr Wissen mit mir zu teilen.

Meine Vorkenntnisse zum Thema künstliche Intelligenz

Mit der Entwicklung von künstlicher Intelligenz hatte ich in der Vergangenheit wenig zu tun. Ich hatte zudem immer ethische Bedenken zu der Entwicklung von künstlicher Intelligenz gehabt. Neben dem Wissen, das einem aus dem Internet, wie zum Beispiel aus dem Nachrichtenportal Tagesschau.de so zufließt, wusste ich, dass googles Suchmaschinenalgorithmus auch irgendwie auf einer KI beruhte. Zudem hatte ich mir einige Bilder des google-Dreams Projektes angeschaut, bei denen mithilfe einer KI Bilder künstlerisch animiert oder angepasst wurden. Auch hatte ich mich mit dem wunderbaren kleinen google-Testdemonstrationsprogramm für Neuronale Netzwerke auseinandergesetzt. Bei einem neuronalen Netzwerk wurden die Funktion und das Zusammenwirken von Neuronen, also Nervenzellen simuliert, um einem Computersystem die Fähigkeit zu verleihen, einen ähnlichen Lernprozess, wie das menschliche Gehirn durchzumachen. Dieses Programm machte die Möglichkeiten von neuronalen Netzwerken, die man zudem selbst über Parameter optimieren konnte, sehr anschaulich.

Auch hatte ich mir die Videos angeschaut in denen die DeepMind-KI Alphastar gegen den Starcraft 2 Pro Gamer MaNa vom Team Liquid spielte und die unter bestimmten Umständen, nämlich, wenn man ihr die Beschränkung des Sichtfeldes der Starcraftlandschaft auf die Größe eines Monitors erließ, scheinbar unbesiegbar wurde. Auch TLO war auf diese Weise besiegt worden. Zugegebenermaßen war das dann aber auch nicht mehr ganz fair, da menschliche Spieler naturgemäß an einen Monitor gebunden gewesen waren. ;-) Ein PC-Welt Artikel, den ich bei google gefunden hatte, gab jedoch an,

dass die KI inzwischen besser war, als 99,8% der menschlichen Spieler und das, obwohl Sie inzwischen der Beschränkung des Sichtfeldes unterlag und auch ihre Möglichkeit der Handlungsgeschwindigkeit, der sogenannten Actions per Minute nach wie vor eingeschränkt gewesen war.

<https://www.pcwelt.de/article/1181733/google-ki-schlaegt-99-8-prozent-aller-starcraft-2-spieler.html>

Danksagung

Ich danke meinem Vater dafür, dass er mir in recht jungen Jahren ermöglicht hat, mich eingehender mit seinem Computer vertraut zu machen und somit erste Erfahrungen mit Computern zu sammeln. Weiterhin möchte ich meiner Mutter danken, für die Liebe und Unterstützung die sie mir über all die Jahre hat zukommen lassen, ebenso wie dem erweiterten Kreis meiner Familie und meinen Freunden, dafür, dass sie für mich da sind.

Im Weiteren danke ich den Bildungseinrichtungen.

Ich bin nun in der dritten Generation von Ingenieuren in der Familie und so hat diese mir mit Sicherheit den ein oder anderen nützlichen Winkeltzug mitgegeben, den ich zum Verfassen dieses Buches oder zu meiner Entwicklung dahin nutzen konnte.

Auch möchte ich den Ärzten, Apothekern, Pflegekräften und Therapeuten, sowie meinen Optikern dafür danken, dass sie im Laufe meines Lebens für mein physisches und psychisches Wohlergehen gesorgt haben.

Ich möchte den Programmierern von Microsoft Word danken, dass ich ihr Programm nutzen konnte, um dieses Buch zu verfassen. Die automatische Rechtschreib- und Grammatikkorrektur war mir beim Schreiben dieses Beitrags eine große Hilfe gewesen.

Weiterhin möchte ich meinem vorhergehenden Arbeitgeber dafür danken, dass ich den Firmenlaptop für meine Weiterbildung zum Thema Elektronik und Informatik nutzen konnte, ebenso, wie meinem gegenwärtigen Arbeitgeber dafür, dass ich auf meinem Firmenrechner diesen Beitrag verfassen konnte. Dafür wäre mein eigenes System eher ungeeignet gewesen, zumal dort nicht einmal Word installiert ist.

Außerdem möchte ich den Mitarbeitern von dict.cc danken. Mit Hilfe dieser nützlichen Seite habe ich in den letzten Jahren so gut wie immer alles nachgeschlagen, was mir aus dem englischen über den Weg gelaufen war und von dem ich die deutsche Übersetzung nicht kannte.

Nicht zuletzt möchte ich der fantastischen weltweiten Open-Source Community dafür danken, dass sie mir Ihrem Engagement diesen Beitrag erst ermöglicht hatte. Einige der herausragendsten Entwicklungen im Softwarebereich waren auf diese Gemeinschaften zurückzuführen, beispielsweise das Linux-Betriebssystem, Firefox, die Wikipedia, die Inhalte von Youtube und Twitch, welche frei zur Verfügung gestellt werden, Teile der Arduino-Programmbibliothek, der Kartenpool der Heroes of Might and Magic Serie und die Open-Source Erweiterungen des Spiels Skyrim, aber vermutlich noch vieles andere.

Selbstverständlich gehört auch das Programm ChatGPT in den Bereich der Open Source Communitys und ich möchte auch dem Tool selbst dafür danken, dass es so kooperativ war und mir akkurate Informationen zur Verfügung gestellt hat.