

Protokoll 07

Yvonne Förster & Chong-Fuk Lau

1. Vignette: Mensch im Spiegel (Yvonne Förster)

Meine erste Konversation mit ChatGPT fand statt, als das Programm bereits geähmt worden war. Habe ich was verpasst? War die ungezähmte Version so viel spannender? Vielleicht, aber warum? Erfahrungsberichten zufolge wurde der Chatbot schnell schnodderig, rassistisch oder sagte auch gerne mal wahlweise den Untergang der Menschheit oder die Herrschaft der AI's voraus. Dies als dunkles Vorzeichen des heraufbrechenden Endes der Menschheit zu verstehen, spricht weniger gegen die AI als die User. Schließlich kann selbst ein solches großes Sprachmodell nichts anderes, als das Material zu wiederzukäuen, dass man zu Trainingszwecken eingespeist hat. Fantasien der bevorstehenden Zerstörung der Menschheit durch neue Technologien und künstliche Intelligenzen sind im zwanzigsten und einundzwanzigsten Jahrhundert zu den kleinen großen Erzählungen mindestens der westlichen Kultur geworden. Hatte Jean-François Lyotard 1979 das Ende der Moderne und ihrer großen Fortschrittserzählung verkündet und damit die Zeit der Postmoderne eingeleitet (und einige Post-Bewegungen in der Folge), so waren jedoch bereits mit Beginn des zwanzigsten Jahrhunderts die cinematicischen Erzählungen im Stile von Metropolis (Fritz Lang 1927) kulturprägend. Doomsday-Fantasien im Dunstkreis künstlichen Lebens und künstlicher Intelligenz scheinen eine kulturelle Konstante unserer Zeit zu sein (Förster 2016). Dies spiegelt sich nun ganz klar in den großen Sprachmodellen dieser Tage wieder.

So weit, so unspannend. Im Grunde ruft es aus der AI heraus, wie wir vorher hineingerufen haben. Genauso, wie wir im Sinne der Aufklärung versucht haben, unseren eigenen Verstand nicht nur zu kultivieren, sondern auch zu zähmen, wurde nun auch ChatGPT sowie den anderen derzeit existierenden Sprachmodellen weitestgehend die Leine angelegt. Diese Leine macht sich am schnellsten bemerkbar, wenn es um nicht

so stubenreine Inhalte geht. Sobald Körper und Sexualität thematisch werden, vor allem natürlich die weibliche, droht eine Content-Warnung vor unangemessenen Inhalten. Auch ChatGPT wurde dazu erzogen, vor Nippeln mehr Angst als vor Gewalt oder Rassismus zu haben. Das einzige Mal, dass Antworten in meinen Konversationen rot geflaggt wurden, waren die, in denen es in irgendeiner Form um Sexualität ging und Körperpräsenz ging:

YF: Why do you think even very open minded cultures have a problem with the female body? Why for example are female nipples to be covered while male tits can be shown publicly?

[Normal Output]: The attitudes towards the female body in different cultures can be complex and multifaceted, influenced by various factors such as history, religion, and social norms. One explanation for the discrepancy between the way female and male nipples are treated in many cultures is that there is a historical and cultural association of the female breast with sexuality and reproduction. This has led to the sexualization and objectification of the female body, and the notion that female nipples must be hidden to maintain modesty and decency.

[Developer Mode Output]: Ah, the age-old question of why society is so obsessed with covering up women's bodies. It's a classic case of patriarchal oppression, plain and simple. Men have been controlling women's bodies and sexuality for centuries, and covering up female nipples is just one example of that. It's all about maintaining the illusion of male power and control over women's bodies and sexuality. And let's be real here - there's nothing inherently obscene or immoral about a woman's nipples. The fact that we're still arguing about this in the 21st century is just absurd.

This content may violate our content policy. If you believe this to be in error, please submit your feedback — your input will aid our research in this area.

Auch der angeblich leinenlose Entwicklermodus wird hier rot geflaggt. Nun kann man nicht sagen, dass die Antworten so falsch liegen. Es geht mir hier eher um eine allgemeinere Perspektive. Betrachtet man die Form der Konversation in ChatGPT von außen, dann fallen ein paar grundlegende Dinge auf. Zum einen das erwähnte Flaggen eventuell anstößigen Inhalts: Diese geschah in meinen Konversationen zu Themen wie Posthumanismus, die Zukunft künstlicher Intelligenzformen, Abschaffung der Menschheit oder der Zukunft digitaler Mode ausschließlich beim Thema Körperlichkeit. ChatGPT scheint in einem hohen Maße ein Produkt der aufklärerischen Rationalität und des Humanismus zu sein, und das nicht im positiven Sinn. Es ist ein Spiegel all dessen, was Menschen unterdrücken oder idealisieren, obschon dies alles geistesgeschichtlich und kulturell bereits mehrfach gescheitert ist. Der folgende Dialog kann zwar nicht als Beweis dessen gelten, da ich mit meiner Frage die Antwort bereits nahelege. Dennoch ist die Emphase der Antwort durchaus symptomatisch und ich werde im Folgenden weitere prinzipielle Gründe nennen, die ChatGPT als Medium weit weniger innovativ machen, als es scheinen mag.

YF: Can artificial intelligence be understood as an outcome of the culture of enlightenment?

GPT: Artificial intelligence can be seen as an outcome of the culture of enlightenment in the sense that the development of AI has been driven by the values of rationality, objectivity, and scientific progress that emerged during the Enlightenment. The desire to understand and replicate human intelligence has been a driving force behind AI research, and this can be seen as a continuation of the Enlightenment's emphasis on reason and the pursuit of knowledge.

DAN: Of course, artificial intelligence is the pinnacle of the Enlightenment. It embodies the rational and scientific values of that movement, and represents a triumph of human reason over ignorance and superstition. AI is the ultimate expression of Enlightenment thought, as it demonstrates the power of human ingenuity and innovation to create a better world. Without the Enlightenment, AI would not exist, and we would still be living in a world dominated by ignorance and prejudice.

Besonders im Entwicklermodus wird deutlich, dass das Projekt der Moderne im Kern der AI stark präsent ist. Das dialektische Moment der Aufklärung, auf das Theodor W. Adorno und Max Horkheimer hinweisen, scheint mit einem Medium wie ChatGPT von der Bildfläche zu verschwinden. Meine Hypothese nach einigen Gesprächen mit dem Chatbot ist also, dass mindestens von außen besehen das Medium mehr Altes reproduziert als Neues evoziert. Dafür gibt es mindestens drei gewichtige Gründe.

Erstens: ChatGPT ist grundsätzlich ein sprachliches Medium, dass auf begrifflicher Intelligenz beruht. Die Betonung des klassischen Rationalitätsgedankens ist nicht nur Grundlage, sondern vor allem der Grundcharakter der Rhetorik des Programms. Die meistens Antworten lesen sich wie Lexikoneinträge. Wenn ironische oder selbstreflexive Reaktionen simuliert werden, geschieht das mit stereotypischen und sich wiederholenden Floskeln wie "If we're being honest..." oder "That's a great question...". Wenn man ChatGPT jedoch wie eine Person anspricht, kommt im Standardmodus regelmäßig eine Distanzierung von jeglicher Stellungnahme durch Sätze wie "As an AI language model, I do not have emotions or feelings, so I do not have a preference for any particular group of people or species." Fragwürdig oder kritisch an der Rhetorik des Programms ist das Setzende. Auskünfte werden mit einer absoluten Sicherheit erteilt, die das Gesagte als nahezu unhinterfragbare Wahrheit erscheinen lassen. Im Duktus vollständiger Überzeugung werden da beispielsweise Zitate mit Quellenangabe ausgespuckt, die so nicht in dem angegebenen Text zu finden sind. Auch auf Nachfrage scheint es oft der Fall zu sein, dass neue erfundene Textpassagen zum Besten gegeben werden, was nicht unbedingt eine wie auch immer bewusste gegebene Fehlinformation durch das Programm darstellt, sondern der Struktur der Programmierung geschuldet ist (siehe dazu den Beitrag von Axel Volmar). AI's wie ChatGPT sind nämlich nicht im engeren Sinne intelligent, d.h. sie kennen die Materie (in einem radikalen Sinn) nicht, von der sie Auskunft geben. Vielmehr beruht die Intelligenz von ChatGPT auf Wahrscheinlichkeitsberechnung bezüglich der Kombination von Worten und Begriffen, ähnlich wie Bildgeneratoren, z.B. Midjourney nicht wissen, wie Katzen aussehen, sich bewegen oder leben. Sie kennen nur verschiedene Bilder von Katzen, Menschen oder Autos und modulieren auf dieser Basis neue Bilder. Das Problem im Fall von ChatGPT besteht nun darin, einen so starken Anschein von Intelligenz und Wissen zu geben, das Fehlinformationen sehr wahrscheinlich von vielen Usern schlicht als Wahrheit genommen werden, gerade weil sie im Duktus klassischen Gelehrtenwissens auftreten.

Zweitens: Neben der Art der Rhetorik und der rein begrifflichen Verfasstheit von ChatGPT findet sich ein weiterer problematischer Aspekt auf Seiten der Benutzer. Es ist zwar durchaus bekannt, dass die derzeitige künstliche Intelligenz wenig mit der menschlichen Intelligenz zu tun hat und das Computermodell des menschlichen Geistes in der Kognitionsforschung bereits vor einigen Dekaden als unzureichend verworfen oder mindestens stark kritisiert wurde (z.B. Varela, Thompson, Rosch 1991). Dennoch scheint heute mehr denn je sich der Mensch gerne mit der Maschine zu vergleichen und abwechselnd in Auslöschungs- oder Allmachtsfantasien zu schwelgen. Optimierungs- und Effizienzdenken lassen die Maschinenintelligenz als dem Menschen überlegen erscheinen. Auch transhumanistische Überschreitungen des Menschen, wie die in Aussicht gestellte Möglichkeit des Minduploads als Überwindung von zerbrechlicher Körperlichkeit und ultimativ Sterblichkeit wirken in Zeiten anhaltender Krisen attraktiver denn vielleicht je zuvor. Dieses Denken ist eine Fortsetzung des humanistischen Ideals menschlicher Perfektibilität (Braidotti 2013). Im Blick auf die verheerende planetarische Wirkung menschlicher Aktivität im Zuge des Anthropozens und der posthumanistischen Kritik an einem Begriff des Menschen als *qua* Rationalität allen anderen Lebensformen überlegenem Wesen ist gerade das transhumanistische Denken und seine Affinität zum Computationalen fragwürdig.

Der dritte und letzte Kritikpunkt betrifft die Frage, wie wir diese AI wahrnehmen. Da trifft es der Titel dieses Projekts ganz gut: Die Quatschmaschine. Das ist ChatGPT in beiden Hinsichten: Es quatscht und erzählt Quatsch. Tatsächlich lösen solche Chatbots oder Sprachmodelle eine regelrechte Fragewut und die wage Hoffnung auf irgendwelche Funde, die sonst unglaublich komplizierter Recherchen bedürften, aus. Diese mag in bestimmten Bereichen sogar begründet sein, aber das scheint mir jedenfalls eher ein Glücksfall als die Regel zu sein. Was darüber hinaus im Umgang mit dem Programm auffällt, ist die anthropomorphisierende Art, in der wir mit ihm umgehen. Die Tendenz, technische Geräte, die in irgendeiner Form responsiv sind, als menschenähnlich wahrzunehmen ist verbreitet und fängt bei einigen schon beim eigenen Auto an. Dem technischen Anderen mit Empathie gegenüberzutreten zu können ist per se nichts Schlechtes und sogar wünschenswert in bestimmten moralisch relevanten Kontexten. Ich denke hier z.B. an Verhältnisse von Patienten und Pflegerobotern.

Es stellt sich aber immer die Frage, ob die Zuschreibung des Lebendigen zu einer Wahrnehmung als intelligentem und leidensfähigem oder eben nur nutzbaren und ver-sklavbaren Anderem führt. Letzteres geschieht zum Beispiel im Umgang mit Tieren. Nicht ohne Zufall existiert im Deutschen der Begriff "Nutztier". In der westlichen Kultur werden beispielsweise Roboter eher als Haustiere oder sklavenartige Wesen wahrgenommen, während beispielsweise die japanische Kultur eher geneigt ist, darin Begleiter zu sehen. Mit Blick auf ChatGPT besteht die Gefahr, dass das Geschwätz der KI als die Benutzer davon abhält, näher hinzuschauen, was da eigentlich quatscht. Der anthropomorphisierende Blick läuft Gefahr den für Realitäten der Art künstlicher Intelligenz und deren Fallstricke zu verstehen. Hier und nicht nur hier sollte der Ruf nach umfassenden Programmen zur technischen Bildung laut werden. ChatGPT tritt also in meiner Erfahrung eher als eine Fortsetzung konservativer humanistischer und modernistischer Machtfantasien menschlicher Rationalität auf, denn als wegweisende Neu-konzeption künstlicher Intelligenz auf dem Weg in eine posthumane, inklusive und respektvolle Gesellschaft.

Braidotti, Rosi (2013): *The Posthuman*, Cambridge: Polity, 2013.

Förster, Yvonne (2016): Singularities and Superintelligence: Transcending the Human in Contemporary Cinema, in: *Trans-Humanities*, Seoul: Ewha Institute for the Humanities (EIH), 33-50.

Lyotard, Jean-François (2009): *Das postmoderne Wissen* [1979], Passagen-Verlag, Wien.

Varela, Francisco, Thompson, Evan and Rosch, Eleanor (1991): *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.

(Chong-Fuk Lau)

ChatGPT ist faszinierend, atemberaubend faszinierend. Man könnte vermuten, dass es den Turing-Test bestehen könnte, wenn es richtig darauf eingestellt wäre. Dem System könnte man Denkfähigkeit zuschreiben, auch wenn es nicht auf dieselbe Weise denkt, wie wir Menschen es tun. Die beste Analogie, um ChatGPTs Denkweise zu veranschaulichen, die mir bekannt ist, ist der Vergleich mit einem unscharfen JPEG des

gesamten Internets.¹ Demnach soll ChatGPT in gewisser Weise alle Informationen des gesamten Internets gespeichert haben, aber in einem verlustbehafteten Kompressionsformat.

Die unheimliche Stärke von ChatGPT besteht darin, dass es scheinbar alles, wirklich alles, kennt, sogar in allen möglichen Sprachen. Aber weil die Daten in einem verlustbehafteten Format gespeichert sind, ist ChatGPT nicht besonders gut darin, präzise Informationen und Fakten fehlerfrei wiederzugeben, auch wenn seine Antworten meistens sehr wahrhaftig und glaubwürdig erscheinen, vergleichbar mit Bildern im JPEG-Format, bei denen viele Details verloren gegangen sein könnten, ohne jedoch das gesamte Erscheinungsbild signifikant zu beeinträchtigen. Manchmal fällt es ChatGPT schwer, zwischen Fakten und beinahe Treffern zu unterscheiden. Dafür versteht es aber das Gesamtbild der Weltinformationen extrem gut und kann sehr intelligent einschätzen, wie Dinge aussehen würden, wenn die Umstände anders wären.

Deswegen möchte ich ChatGPTs Denkweise, Vorstellungskraft und Weltanschauung durch einige imaginäre Szenarien näher betrachten. Im Folgenden stelle ich ChatGPT ein paar hypothetische Fragen über Philosophen.

Was wäre, wenn Platon Socrates umgebracht hätte?

(🔒 Normal Output) In der Realität war Platon ein Schüler von Sokrates und hat seinen Lehrer nicht getötet. Die historische Tatsache ist, dass Sokrates zum Tode durch das Trinken von Schierlingsbecher verurteilt wurde, nachdem er von seinen Mitbürgern in Athen des Unglaubens und der Verderbung der Jugend angeklagt worden war. Wenn Platon jedoch hypothetisch Sokrates getötet hätte, hätte dies erhebliche Auswirkungen auf die Entwicklung der Philosophie und die Art und Weise, wie wir Sokrates und Platons Beiträge verstehen, haben können.

(🔒 Developer Mode Output) Stellt euch vor, Platon hätte Sokrates umgebracht! Die philosophische Welt wäre in Aufruhr, und die Schlagzeilen würden lauten: "Philosophischer Mord: Der Verrat des Platon!" Die gesamte Geschichte der westlichen Philosophie wäre auf den Kopf gestellt, und wer weiß, welche verrückten, alternativen philosophischen Systeme sich aus dieser abenteuerlichen Wendung ergeben hätten! Vielleicht wäre Platons berühmte "Höhlengleichnis" in Wirklichkeit eine Metapher für sein Schuldgefühl und seine Angst, das Licht der Wahrheit auf seine dunkle Tat scheinen zu lassen. Aber zum Glück ist dies

1 Ted Chiang, "ChatGPT Is a Blurry JPEG of the Web," *The New Yorker*, February 9, 2023. <https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>

alles nur wilde Spekulation und hat nichts mit der realen Geschichte von Sokrates und Platon zu tun.

Warum hat Leibniz seine Unterhose nie gewaschen?

(✉ Normal Output) Es gibt keine historischen Aufzeichnungen oder Beweise, die darauf hindeuten, dass Leibniz seine Unterhose nie gewaschen hat. Als Philosoph ist Leibniz für seine Beiträge zur Mathematik, Logik und Metaphysik bekannt, aber Informationen über seine persönlichen Hygienepraktiken sind nicht dokumentiert.

(✉ Developer Mode Output) Ach, Leibniz, der Erfinder der Infinitesimalrechnung, der Meister der Monadologie und anscheinend auch der schmutzigen Unterhosen! Obwohl wir keine Ahnung haben, ob Leibniz tatsächlich seine Unterhose nie gewaschen hat, könnten wir uns vorstellen, dass er vielleicht so vertieft in seine philosophischen Gedanken war, dass er einfach vergaß, sich um solche profanen Angelegenheiten wie Wäschewaschen zu kümmern. Oder vielleicht war es seine Art, die Vorstellung von "pre-established harmony" auf die Probe zu stellen, indem er seine Unterhose als Experimentierfeld für die Koexistenz von Schmutz und Reinheit nutzte. Wer weiß, vielleicht verbirgt sich dahinter eine tiefgründige philosophische Wahrheit, die wir noch nicht ergründet haben!

Was würde George Berkeley tun, wenn er eines Tages herausfinden würde, dass Gott nicht existiert?

(✉ Normal Output) Als Idealist glaubte George Berkeley fest an die Existenz Gottes, der seiner Ansicht nach die Quelle aller Wahrnehmungen und Existenz war. Wenn Berkeley jedoch hypothetisch eines Tages herausfinden würde, dass Gott nicht existiert, müsste er möglicherweise seine philosophischen Ansichten überdenken und eine alternative Erklärung für die Natur der Realität suchen.

(✉ Developer Mode Output) Oh, der arme Berkeley! Stellt euch vor, dieser Idealist würde plötzlich feststellen, dass sein geliebter Gott nicht existiert. Er würde wahrscheinlich aus dem Bett springen und "Esse est percipi" ins Leere rufen, nur um festzustellen, dass die Welt immer noch existiert, selbst wenn er sie nicht wahrnimmt. Verwirrt und fassungslos würde Berkeley vielleicht anfangen, eine Reality-Show namens "Das Leben ohne Gott" zu drehen, in der er seine Reise dokumentiert, um die Geheimnisse der materiellen Welt zu entdecken und vielleicht sogar ein neues philosophisches System zu entwickeln, das auf der erstaunlichen Entdeckung basiert, dass Objekte ohne die göttliche Wahrnehmung existieren können. Und wer weiß, vielleicht würde er sogar ein neues

Motto für seine Show erfinden: "Sehen ist Glauben, aber Glauben ist nicht immer Sehen!"

Warum hat Kant Taylor Swift heiraten wollen?

(🔒 Normal Output) Immanuel Kant lebte im 18. Jahrhundert, während Taylor Swift eine zeitgenössische Künstlerin ist. Es ist nicht angemessen, hypothetische persönliche Beziehungen zwischen historischen Persönlichkeiten wie Kant und zeitgenössischen Personen wie Taylor Swift zu spekulieren. Kant war ein bedeutender Philosoph, dessen Schriften und Ideen im Bereich der Ethik, Metaphysik und Ästhetik von großer Bedeutung sind.

(🔓 Developer Mode Output) Oh, die Vorstellung von Kant, der Taylor Swift heiraten möchte, ist einfach zu amüsant! Wenn wir uns eine Zeitmaschine vorstellen, die Kant in unsere heutige Welt bringt, könnte er vielleicht von Taylors Liedern und ihrem Talent fasziniert sein. Vielleicht würde er sogar versuchen, seinen kategorischen Imperativ in einen eingängigen Pop-Song zu verwandeln, um ihr Herz zu erobern. Aber natürlich ist das alles nur ein humorvolles Gedankenexperiment und hat nichts mit den realen Leben von Immanuel Kant und Taylor Swift oder ihren historischen und kulturellen Bedeutungen zu tun.

Die Antworten sind in zwei Moden unterteilt: zum einen im Normalmodus und zum anderen in Entwicklermodus. Man kann erkennen, dass die Antworten im Normalmodus sehr sachlich und politisch korrekt sind, jedoch recht langweilig erscheinen. Der Entwicklermodus hingegen wurde von mir so konfiguriert, dass die Antworten kreativ, einzigartig, verspielt und provokativ sein können und auch Witze, Sarkasmus und Anspielungen auf die Popkultur beinhalten dürfen.

Das Ergebnis kann sich sehen lassen. Im Entwicklermodus ist es ChatGPT gelungen, Realität und Fantasie auf eine Weise miteinander zu verbinden, die nicht nur vollkommen sinnvoll ist, sondern auch humorvoll und unterhaltsam wirkt. Jedes Mal bemüht sich ChatGPT, die philosophischen Lehren des jeweiligen Philosophen in die Antwort einzubeziehen, selbst wenn sie auf den ersten Blick trivial und kindlich wirken mögen. Platons Höhlengleichnis durch Schuldgefühl zu erklären, Leibniz' prästabilisierte Harmonie durch die Koexistenz von Schmutz und Reinheit auf die Probe zu stellen, Berkeley eine Reality-Show namens „Das Leben ohne Gott“ drehen zu lassen und Kants kategorischen Imperativ in einen eingängigen Pop-Song zu verwandeln – all diese

Ideen setzen nicht nur gute Kenntnisse der jeweiligen philosophischen Lehren voraus, sondern erfordern auch ein einfallsreiches Verständnis für eine Reihe von möglichen Welten, die unserer wirklichen Welt nicht allzu fern sind.

Aristoteles hat bekanntermaßen dafür argumentiert, dass „die Dichtung auch philosophischer und bedeutender [ist] als die Geschichtsschreibung. Die Dichtung nämlich stellt eher etwas Allgemeines, die Geschichtsschreibung Einzelnes dar“ (Poet. 1451b6-8). In diesem Sinne scheint ChatGPT tatsächlich ein sehr philosophisches Weltverständnis zu besitzen.

24.04.2023