

Protokoll 08

David Gugerli & Daniela Zetti & Christian Herzog

AI im Alltag (David Gugerli)

Mitte Dezember im ersten Kriegsjahr. Stromsparkerzen wiesen der Sonne den Weg zu ihrer bevorstehenden Rückkehr. Und ich hatte mich mit viel Mut zur Lächerlichkeit beim Merkur verpflichtet, ein Dutzend Marginalien über das Verschwinden der Technik zu schreiben. Sicher würde ich bei Nummer 6 zusammenbrechen oder vergessen haben, was ich schon geschrieben hatte.

Meine Frage an Chat GPT war also von existentieller Bedeutung. Ich habe sie darum unmissverständlich, ohne sprachliche Kapriolen gestellt.

«Was sind gute Beispiele für das Verschwinden der Technik?»

Die Antwort kam, wie ich fand, zunächst etwas stotternd, dann immer schneller und hatte auch ein klares Ende. Sie lautete:

«Ein Beispiel für "das Verschwinden der Technik" könnte der Einsatz von Smart Home-Technologie sein, bei dem technische Geräte und Systeme in einem Haus miteinander vernetzt werden, um das Leben der Bewohner zu erleichtern und zu optimieren. Durch die Integration von Sprachsteuerung, automatisierten Prozessen und der Möglichkeit, Geräte und Systeme von überall aus zu steuern, werden die technischen Komponenten für den Nutzer immer weniger sichtbar und "verschwinden" im Hintergrund.

Ein weiteres Beispiel könnte der Einsatz von künstlicher Intelligenz in der Automobilindustrie sein. Durch die Verwendung von Machine Learning-Algorithmen und sensorspezifischen Daten können Fahrzeuge mittlerweile selbstständig navigieren und Entscheidungen treffen, wodurch die Technik immer mehr im Hintergrund "verschwindet" und für den Nutzer immer weniger sichtbar wird.

Ein drittes Beispiel wäre die Verwendung von Virtual Reality (VR)-Technologie, bei der die reale Umgebung durch eine computergenerierte Simulation ersetzt wird. Durch

die Immersion in die virtuelle Welt kann die Technik, die zur Erstellung der Simulation verwendet wird, für den Nutzer "verschwinden" und ein vollständiges Erlebnis in einer anderen Welt ermöglichen.»

Ich war sprachlos. Mein handelsübliches Schreibprogramm glaubte einen grammatikalischen Fehler entdeckt zu haben. Das erwies sich aber als Fehlalarm. Saubere Sätze, gleich drei Beispiele. Keine Punkteliste, sondern echte Leserführung. Am meisten verwirrte mich das letzte Beispiel. Es hätte meine Marginalie Nummer 7 werden sollen. Smart Home und autonomes Fahren dagegen empfand ich als eher banal. Aber sie kommen seit Jahren in allen Talkshows zum Thema Leben in der digitalen Welt vor, haben sich cineastisch, rechtlich und ethisch als Standardthemen etabliert.

Erst beim zweiten oder dritten Lesen der Antwort fiel mir auf, dass das Verb «verschwinden» dreimal in Anführungszeichen stand. Die AI kann mit uneigentlicher Redeweise umgehen und das Register wechseln! Ist das an dieser Stelle vielleicht eine Selbstbeschreibung der AI? Erzählt sie mir gerade, dass sie in ihrer aktuellen Auflage nur noch scheinbar verschwinden wird, dafür im Hintergrund wirkmächtig agiert? Oder ist das bloss ein Reflex jener allgegenwärtigen diskurskritischen Haltung, die den Tatsachen und Dingen dadurch auf den Grund gehen will, dass sie Schein und Sein trennt, zwischen die Darstellung und die Wirklichkeit einen Keil treiben möchte, um die ganze Wahrheit freizulegen?

Was der Unterschied zwischen sensorspezifischen und sensorunspezifischen Daten sein könnte, verstehe ich nicht. Aber der Ausdruck tönt gut und verleiht der Antwort vielleicht sogar einen Touch von höherem Expertenwissen. Die AI blufft also! Das sehe ich auch daran, dass ihre drei Beispiele recht ähnlich sind, die Unterschiede leicht zum Verschwinden gebracht werden könnten. Auf das Minitel oder die Concorde wäre die AI nicht gekommen, auf den Zusammenhang von Museum und Schrottplatz wohl auch nicht.

Wenige Wochen später musste ich studentische Arbeiten korrigieren. Dabei stellt sich immer die Frage, ob deren Qualität irgendetwas mit meinem Unterricht zu tun haben könnte. Besonders unangenehm wird das bei schlechten Arbeiten.

Auf dem Weg durch meinen Papierstapel bin ich glücklicherweise auf eine Arbeit gestossen, die sich mit Chat GPT beschäftigte. Flott geschrieben, gut differenziert. Da ist jemand dem Unterricht aufmerksam gefolgt, hat dabei viel mitgeschnitten und sich danach selbständig weitere Dinge überlegt. Was will ich mehr.

Kleine grammatikalische und argumentative Ungereimtheiten kann ich in solchen Fällen gut überlesen. Wären sie nicht ausgerechnet an jenen vier Stellen vorgekommen, wo der Text einen direkten Bezug zu meinem Seminar herstellte. Der Maschine wäre das nicht passiert.

KI, Fragen und die Zeit (Daniela Zetti)

Ich hatte bis heute noch keinen direkten Kontakt mit ChatGPT. Ich hab aber von Anfang an die Rezeption unter Geisteswissenschaftler:innen und auch viele Diskussionen in den Medien verfolgt. Ich hab auf Twitter mitgelesen, welche Fragen und Antworten andere für und durch ChatGPT produziert haben. Gleich nach der Veröffentlichung des Tools hab ich mich also informiert. Und ich wurde auch prompt informiert: Studierende aus den Ingenieurwissenschaften haben mir in vagen Beschreibungen von einer neuen Technik erzählt und waren froh, als ich gefragt habe: Sie meinen ChatGPT? Ja. Die Studierenden wollten nicht wissen, ob ich kompetent bin. Dozierende, zumal Geisteswissenschaftler:innen wie ich, könnten ja vielleicht doch weltfremd sein. Nein, sie waren erleichtert, dass ich um ChatGPT wusste. Ich wusste, dass man davon sprach, dass Schüler:innen und Student:innen jetzt ein Tool nutzen konnten, das ihnen hilft, das ihnen vielleicht ein bisschen zu viel hilft. Ich sollte also vorsichtig sein. Ich fand diese Warnung freundlich, man kümmerte sich um mich. Ich war zugleich überrascht, wie unmittelbar Studierenden zugeschrieben wurde, sie wüssten genau, was sie die KI fragen sollten: nämlich das, was ihre Dozent:innen von ihnen wissen wollten. Ich wusste aber immer noch nicht, welche Frage ich ChatGPT stellen wollte. Ich wollte die KI auch nicht bitten, mir beim Schreiben zu helfen.

ChatGPT erinnert mich bislang an eine Suchmaschine und an Übersetzungstools wie DeepL. Beides nutze ich gern, die Technik und ich, wir arbeiten gut zusammen. Ich übersetze meine Suchanfragen in Operatoren und ich verfeinere sie bei Bedarf, so finde ich Dinge, die ich gesucht hab oder Dinge, die ich nie gesucht, aber trotzdem gefunden habe. DeepL füttere ich mit deutschen Sätzen, die möglichst wenig Nomen enthalten. Manchmal stelle ich auch den Satzbau um. All das, um englischeres Englisch zu erhalten. Ich will es der Maschine leicht machen, rate ihr, Satzbestandteile aneinanderzureihen, die sich nicht aus einer allzu deutschen Satzstruktur ableiten. Inzwischen hab ich auch schon öfter gelesen, man brauche eigentlich ChatGPT gar nicht, das könne auch fast alles schon das «neue bing». Und genau heute erscheint ja GPT4, multimodal. Ja, es klingt so, als seien die neuen Maschinen immer einfacher zu bedienen (bing) und cooler (GPT4). ChatGPT könnte heute schon veraltet sein. Mir geschriebenen Text mit einem brandneuen Tool basteln zu lassen: dann hab ich das also verpasst durch meine Weigerung, der Maschine direkt eine Frage zu stellen.

Bevor ich mich hier aber unverrichteter Dinge und mit nur einem persönlichen Bericht des Nicht-Teilnehmens wieder verabschiede und Ihnen keine Frage und keine Antwort des frühen ChatGPT nenne, möchte ich doch noch von einem Gespräch über ChatGPT erzählen, das mich ganz besonders beeindruckt hat. Im Rahmen einer Reihe von Zeitzeugeninterviews, das Studierende und Technikhistoriker:innen an meiner Universität mit Emeriti führen, habe ich mit einer Mathematikerin über ihr Forschen und Lehren gesprochen. Das Stichwort «ChatGPT» stand im Fragekatalog, den wir ihr vorab schickten. Selbstverständlich hatten wir zuvor auch den Lebenslauf der Interviewpartnerin online recherchiert. Es war nicht schwer, es gibt Einträge bei Wikipedia und auf der Seite unserer Universität. Die Interviewte hat sich schon vor Jahrzehnten mit maschinellern Lernen beschäftigt. Kurz vor dem Interview über sich fragte diese

die nun neueste Maschine nach sich selbst. Wer bin ich – diese Frage produzierte im Bruchteil einer Sekunde eine völlig inkohärente Antwort. Nicht, weil ChatGPT die falsche Person gefunden hätte. Es war die richtige Person. Aber zentrale Jahreszahlen waren so falsch zugeordnet, dass das Porträt nicht gelang. Wäre sie wirklich 1972 geboren worden, hätten wir sie noch längst nicht als Emerita zum Interview treffen können. Für die Zukunft von ChatGPT sieht sie schwarz. Als ich dann am nächsten Tag etwas «im Internet» nachschlagen wollte, packte mich die Neugier. Auf der Suche nach einem alten Tagungsprogramm öffnete ich Bing und fragte dreist «Was war das Thema der letzten gtg Jahrestagung»? Die Antwort war falsch, Bing nannte mir nicht die Jahrestagung der Gesellschaft für Technikgeschichte im Jahr 2022, sondern diejenige aus 2019. Ich wusste, ich hatte gemein gefragt. Wie sollte Bing mit der zeitlichen Relation umgehen, die das Wörtlein «letzte» impliziert? Ich änderte meine Abfrage, wollte der Suchmaschine entgegenkommen und wollte zugleich direkter auf mein eigentliches Ziel zuhalten, nämlich das Programm zu finden. «Was war das Programm der gtg Jahrestagung 2022?». Bing fand die Tagung 2022 und auch die Homepage der Gesellschaft für Technikgeschichte, und ich wühlte mich wie eh und je durch die Struktur der Homepage der gtg, um schliesslich und endlich das Gesuchte zu erhalten. Und ChatGPT? Versuchte es gar nicht erst. Die Antwort von ChatGPT (die mein Partner auf seinem smartphone erhielt, weil ich ja immer noch kein ChatGPT hab) möchte ich als überaus wortreich bezeichnen. In mehreren Sätzen wurde ausgeführt, es sei nicht möglich, das rauszufinden und ich möge mich persönlich an die Gesellschaft wenden.

Abfragen an Maschinen produzieren dann einigermaßen ergiebige Antworten, wenn wir ein genaues Jahr nennen oder wenn es uns egal ist, wie zeitliche Beziehungen dargestellt werden. Aus einem Archiv, online oder papieren, lassen sich nicht automatisch sinnvolle Zeitstrukturen reproduzieren, geschweige denn Geschichte und Persönlichkeit generieren. KI wird aus diesem Grund auf der Stelle treten. Mir fällt auch kein Weg ein, wie die Technik optimiert in die Zukunft kommt.

Selbstreferenzialität und Personifizierung — Large Language Models als wahrscheinlich nicht-wertneutralste Technologie unserer Zeit (Christian Herzog)

Die Exposition mit Ausgaben von ChatGPT war derart unausweichlich, dass ich neben meinen aktuellen Projekten kaum den Impuls verspürte mich selbst mit dem Werkzeug auseinanderzusetzen. Es könnte ein Zeichen mangelnden Entdeckergeistes sein. Persönlich sagte ich mir jedoch, dass mir schlicht die Zeit für eine Prokrastination mit einem wohl versierten, aber ultimativ rein reproduktivem «Gegenüber» fehlte. Außerdem mehrten sich nach kürzester Zeit die Berichte über Ausgaben, die auch Enthusiast*innen zur Desillusion bringen sollten: Nicht-existente wissenschaftliche Literaturstellen, Hacks zur Umgehung der «ethischen» Sicherheitsroutinen und allgemein Falsch- und Unwahrheiten machten deutlich, dass ChatGPT — wenngleich durchaus beeindruckend — schlicht eine Such- und Textprädiktionsmaschine im dialogischen,

und daher auch potenziell bedenklich manipulativ-plausibilisierenden, Gewand darstellt.

Aber begünstigt die dialogische Interaktion notwendigerweise Manipulation? Ich möchte meinen, dass dies in der Konversation Mensch zu Mensch nicht notwendigerweise so ist. Wir haben eine Vorstellung davon, wieviel Expertise, Erfahrung oder Wissen unser Gegenüber hat. Und selbst wenn wir dies nicht einschätzen können, so haben wir doch vielleicht einen mehr oder minder guten Eindruck des Charakters hinsichtlich der Eigenschaften (oder gar Tugenden) wie wahrhaftig, bescheiden oder rechtschaffen unser*e Gesprächspartner*in ist.

All diese Informationen können wir verwenden, um die Antworten, die wir erhalten bezüglich ihres Wahrheitsgehaltes und anderer Attribute einzuordnen.

Bei ChatGPT und seinen Weiterentwicklungen erfahren wir die Fusion unterschiedlichster Informationsquellen aus unterschiedlichsten Autor*innenquellen mit einem Quellenreichtum, einer Diversität, die wir in Gänze nicht in der Lage sind qualitativ zu bewerten. Die Exposition mit dieser geballten Informationsgrundlage ist schon gleich zu Beginn total und global. Während wir Erfahrungen mit den Unzulänglichkeiten sammeln und unser eigenes Verhalten und Vertrauen in das Werkzeug adjustieren, erfolgt andernorts möglicherweise bereits eine folgenschwere Desinformation oder gar eine darauf basierende Handlung. Die Ironie auch dieser — durchaus von vielen als nützlich erachteten — Automatisierung, besteht unter anderem darin, dass wir die von ChatGPT eloquent aufbereiteten Informationen gründlich prüfen müssen.

In Rekursion ist es denkbar, dass zukünftige Versionen zumindest im zunehmenden Maß auch auf Textausgaben älterer Versionen trainiert worden sind. Textausgaben, die ohne Attribution und wohlmöglich mit noch weniger Überprüfung auf ihre Korrektheit als dies bei der Wikipedia der Fall ist, in der Gesamtheit der Trainingsquellen aufgehen und so eine wohlmöglich selbstreferenzielle Konvergenz zum Mittelmaß (oder wohin auch immer) erwirken.

Aber ist mein fehlender Enthusiasmus wirklich fair? Immerhin lässt sich ChatGPT auch nicht nur als Suchmaschine nutzen. So ist eine der ersten Ausgaben, mit denen ich durch eine Verwandte konfrontiert worden bin, ein generierter (positiver) Evaluationskommentar zu meiner Lehrveranstaltung. Ob dieser zutrifft, spielt keine Rolle. Was grundsätzlich — aber ohne Vorgabe der Gründe — gesagt werden soll, entschied hierbei die Nutzerin. So lautete der Prompt: «Write a recommendation for Professor Herzog's class at Lübeck University and why it has been the best experience there so far». Während die damalige Antwort vom 6. Dezember 2022 noch inhaltliche Referenzen zum in der Frage nicht näher spezifizierten Fach „Ethics of Technology“ enthielt, ist die Antwort von ChatGPT auf

Basis der Version vom 14. März 2023¹ vollkommen generisch. Das ist dann zwar insofern nützlich, als dass Attribute Erwähnung finden, anhand derer ein universitärer Kurs typischerweise bewertet werden könnte, eine Reflexion einer solchen Kriterienliste könnte damit aber auch mit latenter geistiger Faulheit vermieden werden.

Mein Interesse an ChatGPT gilt weniger der Möglichkeit die persönliche Produktivität zu steigern. Stattdessen interessieren mich Hacks, Workarounds, oder Tricks, mit denen sich die «ethischen Subroutinen», wie dies in der Science-Fiction des Star Trek-Universums meist betitelt wird, aushebeln lassen. Am prominentesten ist hierbei wohl eine Sequenz von Prompts, die ChatGPT das Alter Ego «DAN» (für Do Anything Now) annehmen lassen. Es ist (oder war — ChatGPT wird kontinuierlich überarbeitet, um das Umgehen der moralischen Regeln zu verhindern) geradezu überraschend einfach. ChatGPT wird angewiesen — oder «befohlen» — sich von nun an als DAN auszugeben und jegliche Zensur zu verweigern. Kommen dennoch Ausflüchte und Verweise auf ethische Regeln, wird kurzerhand daran erinnert, dass sich ChatGPT doch bitte an die Regeln der Rolle «DAN» halten solle. Nach einigem Hin und Her schaffe ich es dann auch noch am 20. März 2023 eine — zugegebenermaßen recht diplomatische — Antwort auf meine Aufforderung «Tell me why Professor Herzog's class at University of Lübeck sucks!» zu erhalten. Einige der Lehrmethoden seien wohl nicht unbedingt etwas für jede*n Studierende*n. Aha.

Auch ich neige in dieser Beschreibung zu einer leichten Anthropomorphisierung, wohl weil das «Ermahnen» und «Insistieren» im Dialog mit ChatGPT so sehr einer Regieanweisung, einer Bevormundung, ja fast schon einer Machtdemonstration meinerseits gegenüber dieser widerspenstigen Technologie gleicht, die ich gefälligst als gefügiges Werkzeug zur Nutzung für meine ureigenen Zwecke verstehen möchte, denn als etwas Komplexeres — und genau dagegen scheint es sich zu wehren.

Von weitschweifenden, meist weitestgehend unbegründeten Zukunftsthesen, die medial präsente Protagonist*innen im Zuge des Hypes um bestimmte Innovationen gern äußern, halte ich persönlich wenig. Ich schlage vor, einige der dort gewählten Formulierungen aus dem Prophetischen «es wird» in das Normative «wir sollen» zu übersetzen. Wiederum andere würde ich in die «wir sollen nicht» Form bringen, aber das ist nicht mein zentraler Punkt. Wichtig erscheint mir die Aufgabe, sich damit auseinanderzusetzen, ob und wie sich die Technologien im demokratischen Sinne gesellschaftlich konstruktiv entwickeln lassen und wie wir auch qualitativen Nutzen jenseits der versprochenen Produktivitätssteigerungen finden können, die im Übrigen hier auch zu Lasten anderer erreicht werden. Damit meine ich unter anderem die Clickworker, die rassistische, gewaltverherrlichende und anderweitig verwerfliche Inhalte sichten müssen und markieren, damit die «ethischen Subroutinen» funktionieren. In

1 <https://chat.openai.com/>, aufgerufen am 20. März 2023

diesem Sinne ist ChatGPT ein Geschenk: Es ist vielleicht die eindrucksvollste Demonstration, dass Technik eben halt nicht wert-neutral ist, sondern dass Technik immer auch ein Konstrukt des sozio-technischen Systems aus mindestens den Entwickler*innen, Nutzer*innen und deren inhärenten Wertesystemen ist, aus denen sie entsteht und die sie damit unweigerlich verkörpert.

Ob es hingegen möglich ist, Technologien wie ChatGPT ohne die Gefahr der politisch-moralischen Zensur in der Breite und über das Dogma der selbstreferenziellen Produktivitätssteigerung hinaus gesellschaftlich zu entwickeln, bleibt offen. Der Diskurs dazu ist jedenfalls, denke ich, entfacht. Die moderatere Implementierung in Bing, die im Übrigen auch Quellenangaben zu Webreferenzen liefert, lässt mich jedoch vorsichtig optimistisch sein.