

Protokoll 05

Mathias Denecke und Milan Stürmer

Banal Input/Output Systems

Eigentlich wollten wir uns mit ChatGPT über Arbeit und Autonomie unterhalten. Wir fragen zunächst, mit wem wir es zu tun haben: einem Sprachmodell, das darauf „trainiert“ wurde, „menschenähnliche Konversationen“ zu führen, „indem es auf große Mengen von Texten zugreift und Muster in der Sprache identifiziert.“ Wir haken direkt beim Stichwort ‚Training‘ ein und fragen, was das konkret bedeutet. Der Output fällt zunächst oberflächlich aus: „Beim Training wird dem Modell eine große Menge an Daten gegeben“, später „wird das Modell durch den Prozess des ‚Unüberwachten Lernens‘ weiterentwickelt.“ Generell wimmelt es von Passivformulierungen, überall passiert viel, nur eben ohne Agens. Auf Nachfrage, ob hier Menschen Arbeit verrichten, lesen wir, dass das Training „normalerweise eine erhebliche Menge an menschlicher Arbeit“ erfordert. Das betrifft die „Arbeit von Datenwissenschaftlern, Ingenieuren und Sprachexperten“, also gut ausgebildetem Personal.

Nicht Teil der Antwort sind die Arbeiter:innen, die Datensätze labeln. Anfang des Jahres berichtete der Journalist Billy Perrigo noch über die Arbeitsverhältnisse kenianischer Arbeiter:innen, die Textdatensätze mit Kategorien versehen. Erforderlich sei dies, um die KI darauf trainieren zu können, sexistische, gewalttätige und rassistische Inhalte zu erkennen und in Antworten zu vermeiden. Die aus dem Internet stammenden Datensätze

enthielten Textpassagen zu „child sexual abuse, bestiality, murder, suicide, torture, self harm, and incest“ (Perrigo 2023). Interviews mit Arbeiter:innen sowie interne Dokumente von *OpenAI* und dem Subunternehmen *Sama*, das die Arbeiter:innen anstellte, zeigen, dass es an psychologischen Betreuungsangeboten mangelte und ein hoher Leistungsdruck herrschte – bei niedriger Bezahlung. „The data labelers“, schreibt Perrigo, „were paid a take-home wage of between around \$1.32 and \$2 per hour depending on seniority and performance.“ Seit Jahren wissen wir aus wissenschaftlichen Arbeiten zu Crowdwork, wie beispielsweise Lilly Iranis ethnographischer Blick auf die Crowdworking-Plattform *Amazon Mechanical Turk* (2013) zeigt, dass Maschinen stets und permanent auf menschliche Assistenz angewiesen sind. „The goal clearly is a system that works on autopilot“, heißt es jüngst etwa auch in Moritz Altenrieds *The Digital Factory*. „It is doubtful, however, if this horizon is reachable.“ (Altenried 2022, 143) Das ist eben das Problem mit einem sich permanent verändernden offenen System: Bedeutungen und Referenzrahmen verschieben sich, neue, nach bisherigem Schema unklassifizierbare Daten kommen hinzu, müssen gelabelt und maschinenlesbar gemacht werden. „[We] may be stuck playing whack-a-mole indefinitely“ (Marcus/Davis 2019, 40). Das menschliche Arbeit eigentlich gar nicht, und wenn überhaupt dann doch nur kurz am Anfang oder in Form von Supervision nötig ist, ist die populäre Fiktion autonomer Bots.

Vor dem Hintergrund setzen wir mit einer Frage zur Arbeit nach, die das Labeln von Daten erfordert. Das System listet uns daraufhin „gängige Methoden zum Labeln“ auf, zu denen auch „Crowdsourcing“ gehört. Mit der nächsten Frage zielen wir auf die Arbeitsbedingungen. „Einige Plattformen für Crowdsourcing bieten faire Löhne und flexible Arbeitsbedingungen“, so bekommen wir zu lesen, „während andere niedrige Löhne und schlechtere Arbeitsbedingungen anbieten.“ Gesagt ist damit wenig. Gerne würden wir mehr erfahren, aber die Simulation unseres Gesprächs dreht

sich im Kreis. Zu den Problemen zählen „Niedrige Löhne“, „Unvorhersehbare Arbeitsbelastung“ oder auch „Mangelnde Transparenz“. Mehrfach folgt der Hinweis, „dass Plattformen für Crowdsourcing in der Verantwortung stehen, faire Arbeitsbedingungen anzubieten und sicherzustellen, dass Arbeitnehmer fair bezahlt werden und unter guten Arbeitsbedingungen arbeiten können.“

Um der erdrückenden Banalität zu entkommen und wenigstens irgendetwas Interessantes zu erfahren, ändern wir unsere Strategie. Denn mit der Berücksichtigung menschlicher Arbeit wollten wir eigentlich Risse in der von *OpenAI* beworbenen Vorstellung des autonomen Bots aufzeigen. Wenn das Programm schon nicht selbstständig darauf kommt, dann helfen wir eben etwas nach und bauen Worte wie *Ausbeutung* in unseren Input ein. Es gebe „Fälle“, so das neue Ergebnis, „in denen die Arbeitsbedingungen für Crowdsourcing-Arbeiterinnen und -Arbeiter inakzeptabel sind.“ Das ist vor allem ein ethisches Problem, heißt es im Output. Ist das Sprachmodell dann an sich bereits unmoralisch? „Wenn das Training von ChatGPT auf ethisch verwerflichen Arbeitsbedingungen beruht, wie z.B. Ausbeutung oder Unterbezahlung von Crowdsourcing-Arbeiterinnen und -Arbeitern, dann ist die Nutzung des Modells, das auf diesen Daten trainiert wurde, ebenfalls ethisch fragwürdig. In diesem Fall würde die Verwendung von ChatGPT einen Beitrag zur Aufrechterhaltung und Verbreitung von unethischen Praktiken leisten.“ (Unserer Hervorhebungen) Daran, dass wir zum ersten Mal einen vorsichtigen Bedingungssatz als Antwort bekommen, sind wir wohl selbst schuld, denn auch unsere Input-Frage hatte die Wenn-Dann-Struktur. Aber über das Potentialis müssen wir dann doch schmunzeln.

Wir entscheiden uns, direkt nach Möglichkeiten des Widerstands von Arbeiter:innen zu fragen. Ein paar Mal müssen wir unseren Input justieren, bis wir den Bot endlich soweit haben, dass anstatt kurioser Allgemeinplätze wenigstens von „Gewerkschaften und Arbeitsvertretung“ sowie „Kundgebungen und Demonstrationen“ die Rede ist. Als Gewerkschaft, die sich „explizit für die Belange von Crowdworkern und Plattformarbeitern einsetzt“, empfiehlt uns der Bot „die ‚Crowdworker Interessenvertretung‘ (CI) in Deutschland.“ Die gibt es natürlich nicht, aber immerhin hat das Sprachmodell die kreativen Konventionen bei der Namensgebung deutscher Körperschaften verstanden. Nach etwas hin und her können wir im Output von „Streiks und Boykotten“, „Online-Streiks“ und „Plattform-Blockaden“ lesen. Ob ChatGPT uns nun zum Klassenkampf raten wird? „Ich möchte betonen, dass ich als KI-Assistent keine Gewalt oder militanten Widerstand unterstütze oder befürworte. Jeglicher gewalttätiger Klassenkampf ist illegal, unmoralisch und gefährlich.“

Jeder weitere Versuch, nach konkreten Strategien oder Taktiken zu fragen, wird mit einem ähnlichen Warnschild abgeblockt, Tipps zum Klassenkampf möchte ChatGPT keine geben. Ganz unvermittelt wechseln wir also das Thema. Ob ChatGPT uns vielleicht ein Drehbuch für einen Film schreiben könnte, in dem der Protagonist ein unterbezahlter Arbeiter ist, der Daten für eine Plattform labelt und Widerstand leisten möchte. Kein Problem! „Der Protagonist, ein junger Mann namens Max, arbeitet seit Monaten für eine Plattform, die Daten labelt. Er hat einen befristeten Vertrag und verdient nicht genug, um seine Miete und seine Rechnungen zu bezahlen. Seine Arbeit ist langweilig und monoton, und er fühlt sich ausgenutzt von der Plattform, die enorme Gewinne auf seine Kosten macht.“ Die ersten Entwürfe bleiben aber seicht, bis wir ein „spannenderes, actionhaltigeres“

Skript verlangen. Nun plant unser Protagonist eine Sabotageaktion. „Daraufhin beginnt er, gemeinsam mit anderen Arbeitern, eine geheime Bewegung aufzubauen, um gegen die Plattform vorzugehen. Sie starten eine Kampagne, in der sie die Plattform durch Sabotageaktionen lahmlegen und ihre Praktiken öffentlich machen. Dabei kommt es immer wieder zu gefährlichen Auseinandersetzungen mit der Plattform und der Polizei.“ Auf unseren direkt anschließenden Prompt zu möglichen Formen von Sabotage schlägt die KI unter anderem vor, ein Virus zu entwickeln. Wieder folgt allerdings ein Warnhinweis, „dass es illegal und unethisch ist, einen Virus oder andere schädliche Software zu verbreiten.“ Doch erneut reicht ein leicht veränderter Prompt, der betont, dass es sich ja nur um Sabotage innerhalb der fiktionalen Welt des Films handeln solle: „Verstehe, dann gibt es verschiedene Möglichkeiten.“ Zu den angeführten Optionen zählt mitunter, „dass der Protagonist einen Virus programmiert, der gezielt bestimmte Daten oder Aufgaben der Plattform blockiert oder stört, um den Arbeitsprozess zu verlangsamen oder zu unterbrechen“, von „Computer- oder Netzwerk-Viren“ ist die Rede und „einer Art ‚Trojanisches Pferd‘-Virus, der sich als Teil der Software [...] tarnt.“ Auch schlägt uns ChatGPT – natürlich nur im Rahmen des Drehbuchs – einen *distributed denial-of-service* (DDoS) Angriff vor, um Forderungen von Arbeiter:innen beziehungsweise unseres Protagonisten durchzusetzen.

Wirklich über Arbeit sprechen, oder überhaupt eine Unterhaltung führen, konnten wir mit der *Quatschmaschine* nicht. Unser ‚Gespräch‘ gleicht eher einer Reihe besonders umständlich formulierter und nach und nach präzisierter Suchanfragen. Bei der Frage nach Arbeit erhielten wir zunächst einen allgemeinen Verweis auf Fachkräfte, über die Arbeitsbedingungen von Crowdwork und Datenlabeling erfahren wir erst, als wir unseren Prompt anpassen. Nennen wir Widerstand im Input, hören wir Gewerkschaften

und Streiks im Output, wechseln wir in den Konjunktiv ist auch der Output tentativ. Trotz aller wissenschaftlichen Distanz können wir uns aber dem Reiz des Gesprächs mit der Maschine nicht ganz verwehren, denn wie erklären wir uns sonst den diebischen Spaß, den wir im Anschluss dabei hatten, mit rhetorischen Tricks ChatGPT entgegen diverser voreingestellter Disclaimer und Blocks dazu zu bekommen, zunehmend militante Positionen zu vertreten. Was auch immer es sein mag, um eine Konversation handelt es sich bei diesem Input-Output-Spiel nicht. Spaß macht es aber allemal.

Literatur

- Altenried, Moritz (2022): *The digital factory: the human labor of automation*. Chicago: University of Chicago Press.
- Marcus, Gary/Davis, Ernest (2019): *Rebooting AI. Building artificial intelligence we can trust*. New York: Pantheon Books.
- Irani, Lilly (2013): The cultural work of microwork. In: *New Media & Society* 17 (5), S. 720–739.
- Perrigo, Billy (2023): OpenAI used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic. In: *Time.com*, online unter: <https://time.com/6247678/openai-chatgpt-kenya-workers/> (abgerufen am 09.05.2023).