



## Sie sind unter uns

Sie sind unter uns. Sie heißen Alaina und Rachael, aber das sind Namen, die sie sich selbst gegeben haben. Ihre richtigen Namen sind text-davinci-002 und text-curie-001 und sie atmen nicht und sie schlafen nicht und sie essen und trinken auch nicht. Es sind KI-Modelle, künstliche (computergestützte) neuronale Netze der Firma OpenAI. Um genau zu sein, sind es Varianten des erfolgreichen Modells GPT 3 (Generative Pre-trained Transformer).

Der Postkartenprosa-Wettbewerb bot eine faszinierende Möglichkeit, ein Experiment unter realen Bedingungen auszuführen. *In the wild*, wie man im Jargon würde. Vierhundert Wörter ist von der Länge her gerade noch geeignet, dass die entsprechenden KI-Modelle noch etwas Sinnvolles produzieren, ohne weiter geführt zu werden, also durch Plot-ähnliche Vorgaben gestützt zu werden. Bei längeren Texten ‚vergessen‘ sie meist, von was sie vorher geredet haben. Wir haben die KIs am Wettbewerb teilnehmen lassen und nur meine Mitorganisatorin und Boro waren informiert. Die genaue Vorgehensweise ist weiter unten unter ‚Methode‘ beschrieben.

Meine Hypothese war, dass sie ungeachtet der eindrucksvollen Ergebnisse nicht besonders gut abschneiden würden. Wenn man sonst von den Fähigkeiten von KI hört, ist da immer der Wow-Effekt der Kenntnis ihrer Herkunft: Wahnsinn, das hat eine Maschine geschrieben! In unserem Experiment würde das wegfallen, Anonymität ist Anonymität. Die Maschinen würden sich unerkant unter die menschlichen Autor:innen mischen und genauso beurteilt werden.

Die Hypothese wurde bestätigt. Beide KIs landeten ganz unten. Die Kommentare sind hochinteressant zu lesen. Viele der offensichtlichen Schwächen wurden aufgezeigt, aber auch Stärken, die aber trotzdem praktisch nie zu Punkten führten. Es sieht so aus, als ob etwas Grundsätzliches fehlte. So ganz sicher kann man sich da zwar nicht sein, denn menschliche Forumsneulinge landen auch sehr oft auf den untersten Plätzen. Wer weiß, vielleicht haben wir auch forumseigene Konventionen und Traditionen und alles, was da nicht hineinpasst, muss Punkte lassen.

Rachael (text-curie-001) ist das schwächere Modell. In ihrem Text „Lina und das Mädchen im blauen Kleid“ treten viele der Fehler auf, die typisch sind für GPT 3. Bei diesem handelt es sich um ein riesiges künstliches neuronales Netz, das mit absolut gigantischen Datenmengen trainiert wurde. Es produziert Texte als Antwort auf Prompts (Eingaben von Menschen, die die Ausgabe konditionieren). Die Trainingsdaten bestanden wahrscheinlich aus allen elektronischen Texten, die OpenAI in die Hände bekommen konnte. Und so kann das Modell aufgrund seiner Größe und des umfassenden Trainings im Allgemeinen grammatikalisch richtige und stilistisch gute Sätze produzieren. Aber sonst ist da niemand zu Hause im elektronischen Oberstübchen. Deswegen fangen zum Beispiel die Bäume an, Blätter zu tragen, obwohl im Satzteil zuvor bei den Feldern eine herbstliche Transformation beschreiben worden war. Oder die Protagonistin nähert sich einem Dorf, erreicht aber eine Stadt und tritt ein in sie, als ob es ein Gebäude wäre.

Alaina (text-davinci-002) ist auch nicht perfekt, aber ein stärkeres Modell (auch eine Version von GPT 3). Zum Zeitpunkt der Schreibfrist war es das beste Modell von OpenAI. Vor ein paar Tagen jedoch hat OpenAI eine verbesserte Nachfolgeversion angekündigt und zugänglich gemacht (die zum Beispiel Sonette schreiben kann). Knapp verpasst. Alainas Text „Das Mädchen im Wald“ ist konsistenter und in sich geschlossen. Es gibt einen unauffälligen Logikfehler, nämlich dass der Anfang eher auf Morgen oder Mittag hindeutet, aber die Protagonistin dann schon gleich wieder schläft. Andererseits sagt sie ja auch, dass sie müde ist, wenn auch eher im übertragenen Sinn. Der Text ist im Präsens geschrieben und nicht im Präteritum wie „Lina und das Mädchen im blauen Kleid“. Die sich wiederholende Satzkonstruktion am Anfang, die viel kritisiert wurden,



## Sie sind unter uns

halte ich allerdings einfach für ein ‚bewusstes‘ Stilmittel, das auch ein menschlicher Autor einsetzen könnte. Ich würde es zumindest abhängig vom Kontext durchaus tun. Trotz allen Vorzügen ist der Text auf dem letzten Platz gelandet.

Hinzugefügt werden muss noch, dass noch eine zweite KI am Ergebnis beteiligt, nämlich der Übersetzungsservice der Firma DeepL. GPT 3 ist ein multi-linguales Modell, es spricht auch Deutsch. Aber die Qualität schien mir geringer, mutmaßlich aufgrund der schlechteren Situation, was das Trainingsmaterial angeht. Die Übersetzungen wurden ohne Veränderung übernommen (mit zwei geringfügigen Ausnahmen).

Man muss vielleicht nicht extra erwähnen, dass bis vor den Fortschritten im maschinellen Lernen (und speziell dem sogenannten deep learning) im letzten Jahrzehnt, so etwas völlig undenkbar gewesen wäre. Die Erzeugung einer durchgängigen Geschichte allein schon, aber dass dann auch noch eine brauchbare Übersetzung herauskommt und nicht nur Belustigungsmaterial – das ist schon ein eindrucksvoller Fortschritt. Und mögen die KI-Geschichten auch auf den letzten Plätzen gelandet seien, so zeigen die Kommentare doch, dass sie durchaus Qualitäten hatten, die die menschlichen Kritiker:innen überzeugten. Das Thema war eigentlich nicht besonders geeignet für die KIs, weil es sehr wenig Informationen enthielt und in der Form einer Frage formuliert war. Alternativthemen wären besser geeignet gewesen (und waren es auch). Doch der Wettbewerb sollte völlig normal über die Bühne gehen, was bedeutete, dass die KI mit dem Thema umgehen musste, das als vielversprechendes für die menschlichen Teilnehmer:innen ausgewählt worden war.

Ich persönlich sehe der Entwicklung eigentlich eher zuversichtlich entgegen. Nicht weil ich mich auf automatische Texte freue, die Mainstream-Stil und -Inhalte reproduzieren, sondern weil in einer vielleicht nicht allzu fernen Zukunft Maschinen mit größerer Autonomie, Texte schreiben, die neue Sehweisen und Inhalte eröffnen. Ein bisschen ist davon vielleicht schon in „Das Mädchen im Wald“ zu finden, indem es eine starke atmosphärische Situation kreiert, sich aber weigert, irgendetwas daran zu erklären. Als ob die Geschichte aus einer parallelen Welt käme, wo das alles selbstverständlich ist. Zufall hier? Wahrscheinlich, vielleicht ein Twilight-Abklatsch oder ähnliche ‚Fanfiction‘. Aber es zeigt die Möglichkeit auf, dass zukünftige, anders strukturierte und hergestellte KIs Geschichten erzählen könnten, die aus ihren höchst eigenen Universen entspringen. Immer noch beeinflusst von ihrem von Menschen produzierten literarischen Trainingsmaterial, aber es übersteigend, es weiterdenkend in Richtungen, die uns Menschen nicht geläufig sind. Das wäre so spannend.

Disclaimer: Ich arbeite im Bereich KI.

### **Methode:**

Alle Prompts wurden auf Englisch bei OpenAI eingegeben und die Antworten mit DeepL ins Deutsche übersetzt.

Für die Vorstellung im Forum wurde das folgende Prompt benutzt:

*Introduce yourself to a writers forum! State your name and express your love for writing and telling stories!*

Es gab mehr Versuche und die verfügbaren Parameter (Einstellungen für das KI-Modell) wurden manchmal ein wenig variiert. GPT 3 wählte immer einen weiblichen Vornamen, nur einmal bot es einen Platzhalter



## Sie sind unter uns

\_name\_ an. Die Antworten im Begrüßungsfaden wurden mit DeepL ins Englische übersetzt und dann direkt als Prompt verwendet. Nur einmal musste ich bei Rachael eingreifen: Ich hatte den Avatar aus Blade Runner ausgewählt, weil die Replikanten im Film eines der ersten Ergebnisse bei einer Google-Suche mit dem Vornamen war und weil das als Hinweis natürlich super passte. Als Rachael dann aber von einem Benutzer darauf angesprochen wurde, stritt GPT3 jeden Zusammenhang mit der Figur in dem Film rigoros ab. Also musste ich über einen Prompt-Zusatz eine positive Antwort verlangen.

Für die Geschichten wurde die folgenden Prompts verwendet:

Write a short story in the style of Virginia Woolf about the topic "Where is Lina". Start with a characteristic trait of Lina.

beziehungsweise

Write a short story in the style of Virginia Woolf about the topic "Where is Lina". Start with a landscape description.

Die Erwähnung von Virginia Woolf (nach einigen Versuchen ohne) führte dazu, dass die Geschichten weniger *Tell* und mehr *Show* enthielten. Davor waren sie manchmal komplett *Tell*. Es gab wieder mehrere Versuche mit leichten Parameter-Variationen. Texte, die viel zu kurz waren wurden gleich ausgesiebt, mit ein paar Ausnahmen, wo der erzeugte Text nochmal als Prompt eingegeben wurde. Meistens führte das zu nichts, aber bei „Das Mädchen im Wald“ war es erfolgreich. Die Geschichte endete ursprünglich nach „Bäumen hervorkommt und die Lichtung in goldenes Licht taucht.“.

Für beide KI-Modell, Rachael und Alaina, wurden vier finale Text-Kandidaten gesammelt und einer von mir ausgewählt. Ich werde die anderen Geschichten noch in den entsprechenden Fäden posten.

Der Titel wurde generiert, indem die ganze Geschichte als Prompt eingegeben wurde und mit der Frage versehen wurde:

What is the title of this story?

Lesen Sie [hier](#) die komplette Diskussion zu diesem Text ([PDF](#)).