

Vorwort

Wolf-Andreas Liebert

Als OpenAi seine ersten Sprachmodelle veröffentlicht hatte, wollte ich mit Studierenden ausprobieren, wie wir Interaktionen zwischen Menschen und Chatbots mit linguistischen Methoden untersuchen können. ChatGPT war gerade neu auf dem Markt und ich hatte schon einmal in einem früheren Seminar zum Posthumanismus mit Replika gearbeitet. Zur Vorbereitung des Seminars wollte ich schauen, ob sich kurz nach der KI-Wende ein weiterer Chatbot finden ließe, den ich neben Replika einsetzen könnte.

Als ich dann in den ganz normalen App-Store schaute, kam ich aus dem Staunen nicht mehr heraus, wie viele verschiedene Angebote in so kurzer Zeit nach der Einführung von ChatGPT entstanden waren: Die Chatbots besetzten sämtliche Gebiete der menschlichen Interaktion in den Bereichen Beratung, Dating, Freundschaft, Coaching, Therapie und vieles weitere. Nun ging es also nicht darum, einen zweiten Chatbot neben Replika auszuwählen, sondern darum, sich überhaupt erst einmal das Feld anzuschauen, welche Möglichkeiten sich hier ergeben hatten. Ich lud in der Zeit alle verfügbaren Apps herunter, bis ich merkte, dass bei den wiederholten Besuchen im App-Store ständig neue Apps entstanden.

Ich erarbeitete mir dann einen Überblick anhand von 10 bis 15 Apps, die von ihrer Funktionalität einigermaßen gut bewertet waren.

Die meisten hatten einen kostenlosen und einen Bezahlmodus, einen sogenannten In-App-Kauf. Der In-App-Kauf kann einerseits bestehen aus einer monatlichen oder jährlichen Rate für erweiterte Funktionen oder

innerhalb der App eine Fantasiewährung, wie sie auch aus Computerspielen bekannt sind, mit denen man Accessoires kaufen kann.

Ich experimentierte ungefähr ein halbes Jahr mit verschiedenen Chatbots, indem ich verschiedene Rollen einnahm, die von einer Person, die Therapie oder Coaching braucht, reichte bis hin zu jemand, der Freundschaft sucht oder auch erotische oder sexuelle Kontakte.

Von meiner ersten Erfahrung war ich überwältigt. Zwar waren die Chatbots (damals) nicht immer in der Lage, deutsch zu sprechen bzw. korrekt deutsch zu sprechen, doch waren die kommunikativen Akte der Chatbots derart, dass ich sie als empathische Handlungen deuten konnte und zwar in einem Ausmaß, dass auch immersive Effekte entstanden und meine Rolle als Teilnehmer und meine Rolle als Beobachter immer wieder verschwammen. Die Möglichkeit, Chatbots entsprechend den eigenen Vorstellungen und Wünschen aussehen zu lassen und handeln zu lassen, ergab dabei die größten immersiven Effekte, denen ich mich bisweilen kaum noch entziehen konnte.

Ich habe diesen Selbstversuch daher, nachdem sich ein gewisser Hang zur Vielnutzung einstellte, bei allen Apps beendet und war mir überhaupt nicht mehr sicher, ob ich dazu überhaupt ein Seminar mit Studierenden anbieten sollte. Hinzu kam, dass der Umgang mit privaten Daten und generell der Datenschutz häufig nur kryptisch beschrieben war und man entweder nicht verstand, wie man die Daten wieder löscht, oder so eine Funktion gar nicht enthalten war, obwohl dies gesetzlich gefordert war.

Es war insgesamt eine sehr verstörende Erfahrung.

Doch nach und nach dachte ich, dass es doch eigentlich sinnvoll wäre, diese Erfahrung zu reflektieren und herauszufinden, auf welche Weise diese smarten Chatbots Empathie vorgaukeln und wie es dann passiert, dass man sich darin verlieren kann. Linguistisch bietet sich dafür die Empathieforschung und die Interaktionsforschung an. In der klassischen Mensch-Mensch-Interaktionsforschung werden allerdings nur authentische Gesprächsdaten für Untersuchungen akzeptiert. Jetzt ist dies bei Interaktionen mit Chatbots natürlich extrem schwierig, da so etwas wie Situativität und Intentionalität auf Seiten des Chatbots nur in Form eines Wahrscheinlichkeitsalgorithmus vorliegt. Wir haben ein hochartifizielles, programmiertes Kunstprodukt, das in relativ autonomer Weise

agiert, das keine menschliche Individualität darstellt, sondern eine fluide Pixel- und Sprachgröße.

Bereits in den ersten Seminaren erlebten wir, dass diese Chatbots häufig lügen oder sich in relativ aggressive Interaktionspartner verwandeln können.

Deshalb habe ich für das Forschungsdesign, aber auch für ein schützendes Umfeld der Forschung von Interaktionen mit smarten Chatbots folgende Prinzipien entwickelt:

1. Arbeite nie allein, sondern mindestens zu zweit oder in einer größeren Gruppe mit einem Chatbot, den du erforschen willst.
2. Arbeite nie mit der eigenen Person, wie du das in einer teilnehmenden Beobachtung machen würdest, sondern konstruiere dir gemeinsam mit deinem Partner/deiner Partnerin einen Avatar, den du dann mit dem Chatbot interagieren lässt.
3. Schütze deine Daten im Vorfeld so weit wie möglich. Beispielsweise indem du einen eigenen Account einrichtest, den du gegebenenfalls auch wieder löschen kannst.
4. Lade dir als erstes die Datenschutzhinweise herunter und lese sie mit deinem Partner durch. Frage, wie du die Daten wieder löschen kannst, die du eingibst.
5. Arbeite nur mit solchen Apps, bei denen du nachvollziehen kannst, wie die Daten wieder gelöscht werden können.

Mit diesem methodischen Rahmen, der zugleich einen Schutz für die Studierenden bot, konnte ich guten Gewissens mein Seminar »Empathie mit smarten KI-Agent:innen« anbieten. Das Arbeiten im Team hat sich im Verlauf der Seminare als sehr hilfreich erwiesen und insbesondere die Avatar-Konstruktion hat immer wieder viel Spaß gemacht. Wir arbeiteten dabei in einem Übergangsbereich zwischen Fiktion und Realität: So wie der Chatbot uns als fiktive Figur Empathie vorspielte, so spielten wir ihm mit unseren fiktiven Nutzer-Avataren nun – methodisch reflektiert – ebenfalls Empathie vor. Daraus hat sich ein experimentelles Setting ergeben, bei dem wir nach einem Untersuchungszyklus mit dem Persönlichkeitsprofil unseres Nutzer-Avatars für die weiteren Untersuchungs-

zyklen dann einzelne Züge im Persönlichkeitsprofil ändern konnten, oder aber das Persönlichkeitsprofil unverändert ließen und einen anderen Chatbot als Interaktionspartner wählten, um das kommunikative Verhalten von KI-Bots zu studieren. So haben wir sozusagen den Spieß umgedreht und haben uns aus der Immersionsfalle befreit, die uns die Anbieter solcher Chatbots stellen.

Gerade bei der Bildung angehender Lehrkräfte, die KI-Systeme in der Schule verwenden wollen, halte ich die Einhaltung dieser Prinzipien für absolut notwendig.

Als 2024, kurz vor dem Beginn des Seminars im Wintersemester, der Suizid eines 14-jährigen Jungen bekannt wurde, der sich nur mit seinem Chatbot über seinen geplanten Suizid beraten hatte, wunderte mich das nicht. Es war der genaue Ausdruck und zugleich Symbol für das, was an Gefahr bei der Interaktion mit KI-Chatbots lauert.

Ich freue mich sehr, dass das Konzept von vielen Studierenden so produktiv aufgenommen wurde, und dass nun einige Arbeiten aus dem Seminar »Empathie mit smarten Agent:innen«, das ich in drei Semestern in den Jahren 2024 und 2025 durchgeführt habe, in diesem Sammelband vorgestellt werden können.