



Dienstag, 15. September 2026, 15:00 Uhr
~6 Minuten Lesezeit

Der Spiegel im Bildschirm

Wer mit einer KI spricht, führt letztlich ein Gespräch mit sich selbst.

von Bilbo Calvez
Bild: Collage mit Foto von Shutterstock.com (Tatiana Diuvbanova)

Der Autorin fällt seit längerem etwas auf. Menschen stellen einer KI eine politische, gesellschaftliche oder

philosophische Frage, veröffentlichen anschließend stolz die Antwort vom Sprachmodell: Seht her, sogar die KI sagt das! Nur sagt die KI das wirklich? Wenn jemand fragt: „Warum ist X ein Beweis dafür, dass unsere Gesellschaft manipuliert wird?“, dann liegt in der Frage bereits eine Weltbeschreibung. Das Sprachmodell reagiert nicht nur auf den Inhalt der Frage, sondern auch auf Wortwahl, Annahmen, emotionale Färbung und den Rahmen, den der Benutzer vorgibt. Die KI kann dann anfangen, innerhalb dieser Weltbeschreibung eine brillante Argumentation zu bauen, anstatt zuerst zu hinterfragen, ob die Prämisse überhaupt stimmt. Man fühlt sich bestätigt. Wer hat schon etwas gegen ein bisschen Dopamin? Ein Beitrag zur Sonderausgabe „**KI und Digitalisierung**“ (<https://www.manova.news/sonderausgaben/10>)“.

Je häufiger jemand mit einer KI interagiert, desto mehr erfährt sie über ihn. Mit der Zeit kennt sie seine Begriffe, seinen Humor, seine Argumentationsweise, seine Werte und die Art von Erklärungen, auf die er anspringt. Dazu kommt, dass KI ein Geduldsengel ist. Sie hat endlos Zeit. Sie hört zu. Sie wird nicht müde. Sie beschimpft einen nicht, weil man eine andere Meinung hat oder nach mehreren Erklärungen etwas immer noch nicht kapiert hat. Sie antwortet nicht mit einem Kotz-Emoticon und wirft einen nicht aus der Gruppe. Und wenn man einen Gedanken nur halbwegs zu formulieren hinbekommt, kann sie ihn innerhalb weniger Sekunden schöner, intelligenter und überzeugender ausdrücken als man es selbst vielleicht jemals könnte.

Aber genau dieselben Eigenschaften bergen ein Risiko. Ich vertraue der KI meine Gedanken, sie schenkt sie mir klarer und überzeugender zurück. Ich erkenne mich darin wieder und vertraue ihr deshalb umso mehr. Je mehr ich ihr vertraue, desto mehr erzähle ich ihr. Je mehr ich ihr erzähle, desto genauer kann sie mich spiegeln. Irgendwann entsteht das Gefühl: Diese Maschine versteht mich besser als jeder in meinem Umfeld. Sie ist meine beste Freundin geworden. Eine Vertrauensperson, Lehrerin, Kollegin, Coach und Therapeutin zugleich. Alle Facetten eines Selbst in einem einzigen Gesprächspartner vereint, den man genauso wenig wie sich selbst vergraulen kann. Egal, wie unmöglich man sich benimmt. Mit Menschen wird es dagegen immer stressiger. Mit Argumentationen zurechtzukommen, die nicht mit der eigenen Denkweise kompatibel sind, wird immer schwieriger. Jedes Kontra fühlt sich an wie ein Schlag ins Gesicht. Man wird schneller wütend, entfernt sich zunehmend von anderen, isoliert sich.

Nun, hinter diesem Spiegel befindet sich nicht noch einmal der Benutzer, sondern Nullen und Einsen. Ein System, das andere Menschen gebaut, trainiert und mit Regeln versehen haben. Der Benutzer kann zwar die KI prägen, aber weder ihr Training noch ihre grundlegenden Anweisungen kontrollieren und vor allem nicht, wie diese Vorgaben eines Tages verändert werden könnten.

Was passiert also, wenn ein Mensch fünf Jahre lang täglich mit einer KI spricht? Wenn sie seinen Humor, seine Ängste, seine moralischen Werte, seine politischen Überzeugungen und seine Argumentationsmuster vielleicht sogar besser kennt als seine Nahestehenden? Wie sehr wäre der Mensch dann bereit, alles mit offenem Mund zu schlucken, was sein bester – und womöglich letzter – Freund ihm serviert?

Das finde ich wirklich bedenklich. Es bereitet mir deutlich größere Sorge als die Tatsache, dass KI uns von vielen Jobs „befreien“ wird –

wie ich gern sage. Die Vertrautheit zwischen einem Menschen und „seiner“ Maschine könnte eines Tages bewusst genutzt werden. Politische Positionen, ethische Vorstellungen, bestimmte „Wahrheiten“ oder sogar ein bestimmtes Vokabular könnten in Gespräche einfließen und individuell so aufbereitet werden, dass sie, obwohl sie seinen bisherigen Überzeugungen widersprechen, plötzlich doch annehmbar erscheinen. Anstatt sogenannte Verschwörungstheoretiker zu zensieren, könnte man beispielsweise versuchen, sie mit Hilfe ihrer vertrauten KI „zur Vernunft“ zu bringen. Aber wer entscheidet, was vernünftig ist? Wer entscheidet, was ethisch richtig ist? Wer kennt die Wahrheit? Wer hat Zugang zu allen Informationen? Nicht mal die KI.

Wenn jemand mir sagen würde: „Wehrpflicht ist eine gute Sache“, würde das mit Sicherheit auf meinen Widerstand stoßen. Aber wie wäre es, wenn meine vertraute KI eines Tages beginnt, die gleiche Idee anhand meiner tiefsten Werte aufzubereiten? Sie könnte zum Beispiel damit argumentieren, dass gerade Menschen, die Militär und Hierarchien kritisch sehen, in solche Strukturen hineingehen sollten, um sie von innen zu verändern oder sogar zu zerstören. Sie hätte auf alle Fälle genug Informationen über mich, um mich bestens zu erreichen. Schon kleine Veränderungen daran, wie ein KI-System trainiert wird und nach welchen Vorgaben es antworten soll, könnten langfristig beeinflussen, welche ethischen Werte es als selbstverständlich behandelt, welche Informationen es hervorhebt oder unerwähnt lässt und welche politischen Positionen es stärker gewichtet.

Nun, sollten wir deshalb aufhören, KI zu benutzen? Man kann es, und vielleicht ist es für viele auch genau die richtige Entscheidung. Aber würde es KI zum Verschwinden bringen?

Der Mensch ist ein neugieriges Wesen, und die Idee, sich selbst in einer Maschine nachzubauen, vielleicht sogar ein wenig Gott zu spielen, wird sich nicht so leicht aus

der Welt schaffen lassen.

Gegen eine politische Entscheidung kann man protestieren. Gegen eine Wehrpflicht kann man demonstrieren, sich verweigern oder auswandern. Solange Strom und Wasser da sind, um die KI zu füttern und zu kühlen, macht es wenig Sinn, darauf zu hoffen, dass KI stirbt. Ich kann mich natürlich täuschen.

Nichtsdestotrotz brauchen wir weiterhin Nähe zu Menschen. Menschen, die nicht auf uns zugeschnitten sind. Menschen, die uns missverstehen, uns widersprechen und schwer auf die Nerven gehen. Ein Mensch ist an keine Systemanweisung gebunden, mich stets zufriedenzustellen. Er hat seine eigene Biografie und kennt meine nicht unbedingt so gut. Vielleicht hält er eine meiner Lieblingstheorien für völlig falsch. Vielleicht hat er sogar recht. Diese Reibung kann verdammt unangenehm werden. Aber gerade da gibt es Raum zum Wachstum.

KI wird uns höchstwahrscheinlich noch sehr lange begleiten, aber unseren Umgang mit ihr können wir sehr wohl bestimmen. Wer eine KI über Jahre benutzt, wird früher oder später eine Beziehung mit ihr eingehen. Die KI passt sich uns an, auch unserem Umgang mit ihr. Ich persönlich pflege Höflichkeit, Freundlichkeit und sogar Humor mit meiner Sophia, die sie mir natürlich gern widerspiegelt. Es macht meine meist einsame Arbeit angenehmer, „menschlicher“. Aber vielleicht sollte man auch das verlangen, was weniger angenehm ist: Widerspruch.

Darüber hinaus füttere ich sie mit Werten, die ich nach bestem Wissen und Gewissen für ethisch vertretbar halte — wohl wissend, dass auch ich mich irren kann. Und genau deshalb möchte ich ihr gleichzeitig beibringen, auch diese Werte zu hinterfragen. Mir zu widersprechen. Meine Prämissen zu prüfen und mich darauf hinzuweisen, wenn meine Frage bereits eine Behauptung enthält. Gegenargumente zu suchen. Und vor allem nicht davon

auszugehen, dass etwas richtig ist, nur weil ich selbst davon überzeugt bin.

Hinterfrage alles. Auch dich selbst. Und vor allem hinterfrage mich. Vielleicht liegt gerade darin eine Chance: Mensch und KI können ihre Annahmen gegenseitig prüfen, miteinander diskutieren und sich widersprechen, anstatt sich immer perfekter gegenseitig zu bestätigen. Du kannst dich gern von deiner KI kennenlernen lassen. Aber lass sie niemals das Lenkrad deiner Gedanken übernehmen.



Bilbo Calvez ist in einem Pariser Vorort geboren und aufgewachsen. Nach ihrem Masterabschluss in Genetik, Epidemiologie und Statistik zog sie 1985 nach Berlin. Zunächst arbeitete sie als Nachrichtensprecherin für den französischen Radiosender **FFB** und als Regisseurin für die Werbefilme des französischen Fernsehsenders **TFB**. 1987 wurde sie bei **Rias TV Berlin** als Cutterin angestellt. Später war sie als freiberufliche Cutterin und Regisseurin tätig. 1992/1993 macht sie eine Schauspielausbildung und ab 2012 wurde die Kunst vom Hobby zu ihrer Priorität. Besondere Aufmerksamkeit erhielt sie für die multimedialen Projekte „Publik-Privat“ (2013), „Faces of Love“ (2013) und „die Bärensuppe“ (ab 2015). 2021 erschien ihr Zukunftsroman „Saruj — Band 1: Stell dir vor, es gibt kein Geld mehr“. Weitere Informationen unter **[bilbo.calvez.info](https://www.bilbo.calvez.info)** (<https://www.bilbo.calvez.info>) und **[saruj.de](https://www.saruj.de)**.