*Wochenkommentar 16/2024 von Matthias Zehnder*

# Der Eliza-Effekt: Warum wir auf Computer reinfallen

Bild: KEYSTONE/DPA/Joerg Carstensen

*Joseph Weizenbaum, aufgenommen auf einem Wirtschaftskongress am 4. April 2001 in Köln.*

**Wie intelligent sind Computer? Verstehen sie uns? Verstehen sie unsere Gefühle und Sehnsüchte? Davon sind heute viele Menschen überzeugt. Auch und gerade Programmierer. Blake Lemoine zum Beispiel war leitender Ingenieur bei Google – bis die Firma ihn entliess. Der Grund: Lemoine war überzeugt, dass die KI von Google keine seelenlose Maschine mehr sei, sondern ein Wesen mit Gefühlen und einem Bewusstsein. Die meisten Nutzer gehen nicht so weit. Immer mehr Menschen fühlen sich aber von den chattenden KI-Programmen gut verstanden. Manchmal sogar besser als von Menschen. Warum ist das so? Was bringt uns dazu, einer Maschine so viel Verständnis und Gefühl zu attestieren? Die Antwort findet sich in einem Experiment, das der deutsch-amerikanische Computerwissenschaftler Joseph Weizenbaum bereits 1966 durchgeführt hat. Das Programm, das er dafür entwickelte, hiess «Eliza». Die menschliche Vertrauenseeligkeit gegenüber Maschinen heisst seither «Eliza-Effekt».**

Die öffentliche Meinung unterscheidet sich manchmal stark vom praktizierten Verhalten. Wir kennen das etwa von der Diskussion um Alkohol am Steuer. Jeder weiss: Alkohol am Steuer ist kein Kavaliersdelikt. Wir sind uns alle einig: «Wer fährt, trinkt nicht – wer trinkt, fährt nicht.» Die Realität sieht anders aus: Viele Fahrer sind überzeugt, dass sie so kompetente Fahrzeuglenker sind, dass ein Glas ihre Fahrtüchtigkeit nicht beeinträchtigt. Oder auch zwei. Die Folge: die Anzahl Autounfälle unter

Alkoholeinfluss hat in der Schweiz eine Rekordmarke erreicht. Über 4500 Unfälle waren es 2022 – jeder zehnte davon war ein schwerer Unfall mit Schwerverletzten oder Toten. Offensichtlich klaffen die öffentliche geäusserte Meinung und das individuelle Verhalten stark auseinander, wenn es um Alkohol am Steuer geht.

Etwas Ähnliches stelle ich derzeit in Bezug auf den Umgang mit künstlicher Intelligenz fest. In der Öffentlichkeit sind sich Fachleute einig: Im Umgang mit der KI ist Vorsicht geboten. Insbesondere Chatbots sollten Sie nicht zu vertrauensselig nutzen: Weder können Sie darauf vertrauen, dass Ihre Daten korrekt geschützt sind, noch können Sie darauf bauen, dass stimmt, was die Chatbots den lieben langen Tag von sich geben. Ihre Sprache ist zwar grammatikalisch korrekt, sie haben aber keine Ahnung, was sie sagen. Die Realität indes sieht anders aus: Alle Welt nutzt die plappernden Programme und attestiert ihnen schon fast Zauberkräfte. «Werden Ärzte bald durch künstliche Intelligenz ersetzt?» fragt zum Beispiel SRF. «Fremdsprachen lernen? Braucht es bald nicht mehr!», ist die «Basler Zeitung» überzeugt.

## Verführerische Plapperprogramme

Immer häufiger wird die KI auch als Gegenüber für «Gespräche» genutzt. «Gespräche» in Anführungszeichen, weil man einen maschinell generierten Dialog nicht als Gespräch bezeichnen kann. Was die meisten Anwender ja auch wissen. Sie nutzen die Dienste aber trotzdem. Warum ist das so? Was steckt hinter der Vertrauensseligkeit, die wir diesen Maschinen gegenüber an den Tag legen? Warum nur fallen wir auf die eloquenten Plapperprogramme rein?

Die Antwort darauf hat Joseph Weizenbaum schon vor fast 50 Jahren gegeben: «Die Macht der Computer und die Ohnmacht der Vernunft» heisst das Buch, das er 1977 veröffentlicht hat. Weizenbaum war alles andere als ein weltfremder Computerkritiker. Er gilt als einer der Pioniere der Computerentwicklung. Als er das Buch schrieb, war er Informatik-Professor am Massachusetts Institute of Technology MIT in Boston. Weizenbaum weiss also, wovon er spricht. Sein Buch beginnt mit folgenden Sätzen:

*In diesem Buch geht es nur vordergründig um Computer. Im wesentlichen wird der Computer hier lediglich als Vehikel benutzt, bestimmte Ideen vorzutragen, die viel wichtiger als Computer sind.*
(Weizenbaum 1977: S. 9)

## Automatisierung der Psychotherapie

Später sagte Weizenbaum immer wieder, er sei kein Computerkritiker, er sei Gesellschaftskritiker. Er kritisierte also nicht die Maschinen, sondern die Art und Weise wie wir Menschen mit ihnen umgehen. Legendär sind seine Experimente mit einem Programm, das er «ELIZA» nannte. Er hatte das Programm 1966 entwickelt. Es war nichts weniger als der erste funktionierende Chatbot, also das erste Programm, mit dem Menschen sich unterhalten konnten. Natürlich nicht mündlich, sondern nur über die Tastatur, aber immerhin. In seinem Buch beschreibt Weizenbaum, wie ELIZA funktionierte:

*ELIZA war ein Programm, das in der Hauptsache aus allgemeinen Methoden bestand, Sätze und Satzfragmente zu analysieren, sogenannte «Schlüsselworte» in Texten zu lokalisieren, Sätze aus Fragmenten zusammenzu-*

*setzen usw. Es hatte mit anderen Worten keinen eingebauten kontextuellen Rahmen oder einen Gegenstandsbereich. Dieser wurde ihm durch ein «Skript» übermittelt. In gewissem Sinne war ELIZA eine Schauspielerin, die über eine Reihe von Techniken verfügte, aber selbst nichts zu sagen hatte. Das Skript seinerseits bestand aus einer Anzahl von Regeln, die dem Schauspieler erlaubten, über jedes Thema zu improvisieren, das vom Skript vorgesehen war.*

*Das erste umfangreiche Skript, das ich für ELIZA geschrieben hatte, versetzte dieses Programm in die Lage, die Antworten zu parodieren, die von einem an der Methode von Roger orientierten Psychotherapeuten bei der Vorbesprechung mit einem Patienten gegeben wurden.*

(Weizenbaum 1977: S. 250f.)

Weizenbaum hat also ELIZA über ein Skript eine Rolle zugewiesen. Wenn der Benutzer etwas eingab, hat das Programm in seinem Skript lediglich nachgeschlagen, was es darauf antworten muss. Die erste Rolle, die ELIZA dabei einnahm, war ironischerweise die Rolle eines Psychotherapeuten. Weizenbaum erklärt in seinem Buch, warum:

*Ein solcher Therapeut ist verhältnismässig leicht zu imitieren, da ein Grossteil seiner Technik darin besteht, den Patienten dadurch zum Sprechen zu bringen, dass diesem seine eigenen Äusserungen wie bei einem Echo zurückgegeben werden.*

(Weizenbaum 1977: S. 15.)

Ein Beispiel bietet der folgende Dialog zwischen einer jungen Frau und ELIZA:

*Junge Frau: Die Männer sind alle gleich.*
ELIZA: In welcher Weise?
*Junge Frau: Sie machen uns immer wegen irgendwas verrückt.*
ELIZA: Können Sie ein Beispiel nennen
*Junge Frau: Na ja, die Idee zu diesem Gespräch stammt von meinem Freund.*
ELIZA: Die Idee zu diesem Gespräch stammt von Ihrem Freund
*Junge Frau: Er sagt, daß ich oft deprimiert bin.*
ELIZA: Es tut mir leid von Ihnen zu hören dass Sie deprimiert sind
*Junge Frau: Das stimmt, ich bin unglücklich.*
ELIZA: Glauben Sie dass dieses Gespräch Ihnen hilft nicht mehr unglücklich zu sein
*Junge Frau: Ich brauche jedenfalls Hilfe, das steht fest.*
ELIZA: Was würde es für Sie bedeuten Hilfe zu bekommen
*Junge Frau: Vielleicht könnte ich lernen, mit meiner Mutter auszukommen.*
ELIZA: Erzählen Sie mir mehr über Ihre Familie

(Weizenbaum 1977: S. 15f.)

Die letzte Wendung im Gespräch ist eine für ELIZA typische Ersetzung: Im Skript steht die Anweisung, nach der Familie zu Fragen, wenn das Gespräch auf die Mutter kommt. Das funktioniert in diesem engen Rahmen, den ein Dialog mit einem Psychotherapeuten steckt. Es sieht auf den ersten Blick aus, als würde ELIZA die junge Frau verstehen. Die Frau sagt, dass sie lernen könnte, mir ihrer Mutter auszukommen. ELIZA antwortet darauf: *Erzählen Sie mir mehr über Ihre Familie.* Genau so würde ein Psychotherapeut reagieren. ELIZA versteht aber nicht, was die Frau sagt. Auch wenn sie einen Satz eingeben würde wie: *Mahalia Jackson ist*

*die Mutter aller Soul-Sängerinnen.* Oder: *Zu dieser Schraube fehlt mir eine passende Mutter.* würde ELIZA darauf antworten: *Erzählen Sie mir mehr über Ihre Familie.* Das Programm reagiert also quasi blind auf das Schlüsselwort «Mutter».

## Gesprächsparodie als echt angenommen

Für Weizenbaum war klar, dass sein Programm eine Unterhaltung nur imitierte, ja parodierte. Er war deshalb überrascht, wie enthusiastisch die Benutzer auf das Programm reagierten. Auch wenn er ihnen die Funktionsweise erklärte, bestanden sie nach der Unterhaltung mit ELIZA darauf, dass die Maschine sie wirklich verstanden habe. Weizenbaum stellte bestürzt fest, wie schnell und wie intensiv Personen, die sich mit dem Programm unterhielten, eine emotionale Beziehung zum Computer herstellten und wie sie ihm eindeutig menschliche Eigenschaften zuschrieben. Es gab praktizierende Psychiater, die im Ernst glaubten, das Computerprogramm könne zu einer automatischen Form der Psychotherapie ausgebaut werden.

Weizenbaum gelang es, mit ELIZA zu zeigen, dass sich eine Kommunikation innerhalb eines bestimmten Rahmens simulieren lässt, weil die Menschen das, was sie hören oder lesen, immer innerhalb eines Erwartungsrahmens interpretieren. Diese Erwartungen, das Bild, das die Menschen von einer Kommunikationssituation im Kopf haben, prägen die Interpretation stark. Weizenbaum schreibt:

*Jeder Gesprächspartner bringt in die Unterhaltung ein Bild darüber ein, wer der andere ist. Da es keinem Menschen möglich ist, den anderen vollständig zu kennen, besteht dieses Bild zum Teil aus Attributen, die der Identität des anderen zugeschrieben werden, die jedoch notwendig auf Anhaltspunkten beruhen müssen, die aus der jeweiligen Lebenserfahrung der beiden Partner stammen. Unsere Erkenntnis einer anderen Person ist somit ein Akt der Induktion aus bestimmtem Material, das uns zum Teil von ihr und zum Teil von unserer Rekonstruktion der übrigen Welt angeboten wird; sie ist eine Art Verallgemeinerung. Mit anderen Worten, wir alle tragen gegenseitige Vorurteile mit uns herum. Und wie wir bemerkt haben, fällt es uns allen schwer oder es ist uns sogar nahezu unmöglich, Beweise wahrzunehmen – geschweige denn zu akzeptieren und sie auf uns wirken zu lassen –, die unseren Urteilen widersprechen. Unter diesen Umständen ist es leicht zu verstehen, warum Menschen, die mit ELIZA eine Unterhaltung führen, den Glauben hegen, dass sie verstanden werden, ohne dass man sie davon abbringen kann. Der «Sinn» und die Kontinuität, die die mit ELIZA sprechende Person wahrnimmt, werden weitgehend von dieser selbst hergestellt. Von ihr gehen die Bedeutungen und Interpretationen dessen aus, was ELIZA «sagt», die ihre ursprüngliche Hypothese bestätigen, dass das System mit Verständnis begabt sei.*

(Weizenbaum 1977: S. 253)

Das menschliche Gehirn ist ein überaus mächtiger Interpretationsapparat, der darauf trainiert ist, auf ganz bestimmte Weise Sinn und Bedeutung in eine Kommunikation zu lesen. Das ist der entscheidende Punkt: Wir Menschen nehmen die Welt nicht neutral und unvoreingenommen auf. Das wäre extrem aufwändig. Stattdessen arbeite wir arbeiten mit Annahmen und Rahmenbedingungen. Wenn wir früher, als die Menschen ihre Zugtickets noch am Billetschalter kauften, an einen solchen Schalter traten und sagten: «Zürich retour», dann wusste der Mann oder die Frau

hinter der gelochten Glasscheibe, was gemeint ist, weil man normalerweise an einem solchen Schalter ein Zugticket kauft. Der Schalterbeamte und sein Kunde bewegen sich in einem engen Erwartungsrahmen. Das macht die Kommunikation effizient.

## Der Mann mit dem Hammer

Solche Erwartungsrahmen haben wir über lange Zeit für den Umgang miteinander entwickelt. Sie vereinfachen es, das Gegenüber ohne viel Aufwand zu verstehen. Allerdings nur dann, wenn das Gegenüber denselben Erwartungsrahmen hat. Im Ausland, im Kontakt mit Menschen aus anderen Kulturen, kann es deshalb zu groben Missverständnissen kommen. Aber auch im alltäglichen Umgang kann es sein, dass unsere Erwartungen sich selbstständig machen. Vielleicht kennen Sie die Geschichte vom Mann, der einen Hammer ausleihen wollte. Der österreichisch-amerikanische Psychotherapeut und Kommunikationswissenschaftler Paul Watzlawick erzählt sie in seinem Buch «Anleitung zum Unglücklichsein»:

*Ein Mann will ein Bild aufhängen. Den Nagel hat er, nicht aber den Hammer. Der Nachbar hat einen. Also beschliesst unser Mann, hinüberzugehen und ihn auszuborgen. Doch da kommt ihm ein Zweifel: Was, wenn der Nachbar mir den Hammer nicht leihen will? Gestern schon grüsste er mich nur so flüchtig. Vielleicht war er in Eile. Aber vielleicht war die Eile nur vorgeschützt, und er hat etwas gegen mich. Und was? Ich habe ihm nichts angetan; der bildet sich da etwas ein. Wenn jemand von mir ein Werkzeug borgen wollte, ich gäbe es ihm sofort. Und warum er nicht? Wie kann man einem Mitmenschen einen so einfachen Gefallen abschlagen? Leute wie dieser Kerl vergiften einem das Leben. Und dann bildet er sich noch ein, ich sei auf ihn angewiesen. Bloss weil er einen Hammer hat. Jetzt reicht's mir wirklich. – Und so stürmt er hinüber, läutet, der Nachbar öffnet, doch noch bevor er «Guten Tag» sagen kann, schreit ihn unser Mann an: «Behalten Sie sich Ihren Hammer, Sie Rüpel!»*

(Watzlawick 1983: S.37-38)

Watzlawick überzeichnet hier, wie Erwartungen und Phantasien die Kommunikation prägen. Dasselbe Phänomen prägt unserem Umgang mit künstlich intelligenten Systemen: Die Menschen haben Respekt vor der Mächtigkeit, ja der Allmächtigkeit des Computers. Das drückt sich ja schon in der Bezeichnung der Systeme als «künstlich *intelligent*» aus. Weil die Erwartung die Wahrnehmung prägt, erleben wir die Systeme auch als intelligent. Wir lassen uns von ihnen verführen. Joseph Weizenbaum hat schon in den 60er-Jahren mit seinem Programm Eliza gezeigt, wie wenig es dazu braucht, dass sich Menschen täuschen lassen. Schuld daran sind nicht übermächtige Maschinen, sondern die Menschen selbst: Es ist der Interpretationsapparat in unserem Kopf. Joseph Weizenbaum spricht deshalb vom «Eliza-Effekt».

### Erwartungseffekt auch in der Schule

Der «Eliza-Effekt» ist ein Erwartungseffekt: Wir nehmen das, was wir erleben, nicht neutral wahr, es ist geprägt durch unsere Erwartungen. Aus der pädagogischen Psychologie kennen wir den Pygmalion-Effekt. Er be-

sagt, dass es einen Zusammenhang zwischen den Erwartungen von Lehrpersonen und den Leistungen der Schüler gibt. Erwartet ein Lehrer von einem Schüler gute Leistungen, zeigt der Schüler im Laufe der Zeit tatsächlich bessere Leistungen und zwar, das ist das Spannende daran, unabhängig davon, ob die Erwartungen gerechtfertigt sind oder nicht. Denn Lehrpersonen schenken Schülern mehr positive Aufmerksamkeit, wenn sie viel von ihnen erwarten, sie fordern und fördern sie eher. Schüler, denen Lehrpersonen positive Erwartungen entgegenbringen, entwickeln ein höheres Selbstvertrauen und sind motivierter, sich anzustrengen.

Hohe Erwartungen an Menschen zahlen sich also aus. Wer jemandem Vertrauen schenkt, wird oft belohnt. Denn die Erwartungshaltung, die ich einem Menschen gegenüber an den Tag lege, beeinflusst sein Verhalten. Im Umgang mit Maschinen wird diese Eigenschaft uns Menschen aber zum Verhängnis: Die positive Erwartungshaltung und die Fähigkeit und Bereitschaft, Sinn und Bedeutung in eine Antwort zu lesen, führt dazu, dass wir Computer hoffnungslos überschätzen. Denn die wissen nicht, was sie tun. Sich von einem Chatbot beraten zu lassen, ist wie ein Gespräch mit einem Papagei. Der mag vielleicht hilfreiche Börsentipps geben, als Psychologe oder als Philosoph ist ein Papagei aber nicht zu gebrauchen.

Herausgefunden hat das alles schon Joseph Weizenbaum in den 60er-Jahren. Es ist deshalb Zeit, sein Buch «Die Macht der Computer und die Ohnmacht der Vernunft» wieder hervorzunehmen. Was die Menschen und vor allem die Medien heute über die künstliche Intelligenz sagen, das lässt sich mit einem Wort erklären: Eliza-Effekt. Denken Sie dran, wenn Sie das nächste mal mit einer KI zu tun haben: Das, was Sie da als Intelligenz wahrnehmen, das ist nicht die Maschine, das sind Sie selbst.

Basel, 19. April 2024, Matthias Zehnder **mz@matthiaszehnder.ch**

Wenn Sie den Wochenkommentar unterstützen möchten, finden Sie unter **https://www.matthiaszehnder.ch/unterstuetzen/** die entsprechenden Möglichkeiten – digital und analog.

## Quellen

Benini, Sandro (2024): *Fremdsprachen lernen? Braucht es bald nicht mehr!* In: Basler Zeitung. [https://www.bazonline.ch/kuenstliche-intelligenz-fremdsprachen-lernen-ist-bald-passe-598312448467; 18.4.2024].

Engel, Andreas (2023): *Anzahl Alkoholunfälle erreicht Rekordmarke*. In: Blick, 31. 12. 2023. [https://www.blick.ch/auto/news_n_trends/ein-kanton-ist-trauriger-spitzenreiter-anzahl-alkoholunfaelle-erreicht-rekordmarke-id19258822.html; 19.4.2024].

Ibrahim, Nadine (2023): *KI in der Medizin - Werden Ärzte bald durch künstliche Intelligenz ersetzt?* In: Schweizer Radio Und Fernsehen (SRF).

[https://www.srf.ch/news/schweiz/ki-in-der-medizin-werden-aerzte-bald-durch-kuenstliche-intelligenz-ersetzt; 18.4.2024].

Merschmann, Helmut (2008a): *Joseph Weizenbaum: Der zornige alte Mann der Informatik*. In: DER SPIEGEL. [https://www.spiegel.de/netzwelt/tech/joseph-weizenbaum-der-zornige-alte-mann-der-informatik-a-527122.html?sara_ref=re-so-app-sh; 19.4.2024].

Merschmann, Helmut (2008b): *Zum Tode von Joseph Weizenbaum: Der Kritiker geht, die Kritik bleibt bestehen*. In: DER SPIEGEL. [https://www.spiegel.de/netzwelt/web/zum-tode-von-joseph-weizenbaum-der-kritiker-geht-die-kritik-bleibt-bestehen-a-540088.html?sara_ref=re-so-app-sh; 19.4.2024].

Nezik, Ann-Kathrin (2023): *Künstliche Intelligenz: Hast du ein Bewusstsein? Ich denke schon, antwortet der Rechner*. In: ZEIT ONLINE. [https://www.zeit.de/2023/03/ki-leben-chatbot-gefuehle-bewusstsein-blake-lemoine/komplettansicht; 18.4.2024].

Watzlawick, Paul (1983): *Anleitung zum Unglücklichsein.* Piper: München 1983

Weizenbaum, Joseph (1977): Die Macht der Computer und die Ohnmacht der Vernunft. Frankfurt am Main: Suhrkamp 1977 (als stw 274)