

KI als Spiegel

Essay von Philippe Wampfler

Je menschenähnlicher Figuren oder Roboter werden, desto vertrauter werden sie uns. Wir empfinden Empathie und fühlen uns in ihrer Gegenwart wohl. Kurz bevor wir sie fast nicht von echten Menschen unterscheiden können tritt ein paradoxer Effekt ein: Sie werden uns unheimlich. Dieser Effekt hat einen Namen, er wird als ‚uncanny valley‘¹ bezeichnet. Die Vorstellung gruselt uns, etwas Künstliches könnte genau so sein wie wir, die wir geboren wurden und gelebt haben.

Als Gesellschaft stehen wir in Bezug auf die KI-Technologie vor der Frage, ob dieses Tal des Unheimlichen, wie das ‚uncanny valley‘ auf Deutsch heißt, durchschritten werden kann. Können wir uns Figuren oder Intelligenzen vorstellen, die so menschenähnlich sind, dass wir das als angenehm empfinden? Der Turing-Test² präsentiert ein Verfahren, um das zu überprüfen: Eine Person erhält auf ihre Fragen jeweils zwei Antworten – eine von einem Mensch, die andere von einer Maschine. Die Maschine besteht den Turing-Test, wenn die Person sie für den menschlichen Gesprächspartner hält. Das Kriterium, um Maschinen echte Intelligenz oder Denkvermögen zuzuschreiben, wäre also: Kann sie ein so lebendiges Gespräch führen, dass wir Menschen gar nicht mehr unterscheiden können, ob wir mit einem Menschen oder einem Automaten sprechen? (Eine Untersuchung aus dem Dezember 2023 zeigt, dass ChatGPT den Turing-Test noch nicht besteht – nur menschliche Nutzer werden aktuell in über 50% der Versuche auch als Menschen identifiziert.)

Der Turing-Test macht deutlich, was an der Vorstellung unheimlich ist, dass Maschinen Menschen in ihrem Handeln gleichen: Wir sind dann nicht mehr in der Lage einzuschätzen, wer ein Mensch ist und wer nicht. Wir lassen uns auf bisher menschliche Formen der Interaktion ein, indem wir etwa ein Gespräch führen – und unterhalten uns mit einer Maschine. Snapchat stellt allen Usern einen Kontakt namens ‚MyAI‘ zur Verfügung: Der Bot lernt einen kennen und ist immer für einen anregenden Chat zu haben. Der Spielzeughersteller Curio integriert diese Technologie in Plüschtiere³, die mit kleinen Kindern sprechen und sich an sie gewöhnen. Was zunächst wie ein nettes Feature wirkt, wird doch schnell wieder unheimlich: Möchten wir Menschen, die einem Bot sprechen, weil der immer positiv und zuverlässig antwortet? Der Technikkritiker Jaron Lanier hat den Turing-Test deshalb auch nachvollziehbar kritisiert: Er zeige nur, dass wir bereit seien, die Ansprüche an menschliche Interaktionen zu senken, wenn wir uns zu viel mit Maschinen abgäben.⁴

Daraus lässt sich eine tiefere Einsicht ableiten, die hinter dem Problem des ‚uncanny valley‘ steht: Warum wirkt es eigentlich unheimlich auf uns, wenn eine Figur uns gleicht? Wenn wir ausgehend davon über die Fragen nachdenken, die uns in der Auseinandersetzung mit KI beschäftigen, eröffnet das neue Perspektiven. Wir können uns fragen, warum es uns beunruhigt, wenn eine Maschine Entscheidungen fällt, moralische Abwägungen trifft oder Sachverhalte so zurechtlegt, dass es uns schwer fällt, Wahrheit, Lüge oder Manipulation zu unterscheiden. Mittlerweile werden die Maschinen sogar faul; Ende 2023 haben sich Klagen gehäuft, KI-Tools wie ChatGPT würden Anforderungen zunehmend kurz und schnippisch beantworten.

Das alles wirkt so, als blickten wir in einen Spiegel. Wir verwenden ein Tool, das scheinbar so viel kann, dem wir aber doch nicht trauen. Eine Maschine, die leistungsfähig aber fehleranfällig ist. Automaten, deren Rechenkapazitäten uns faszinieren, die aber plötzlich die Arbeit verweigern. Sind sie nicht äußerst menschlich? Ist das nicht, was uns als Menschen definiert: Kreativität, Leistungsbereitschaft, Effizienz – aber auch Fehler, Lügen, Faulheit?

Auf einer technischen Ebene überrascht das nicht: Die Chat-KI-Tools sind Datenbanken, die menschliche Texte durchsuchen, Muster identifizieren und basierend darauf Texte oder Bilder hervorbringen, die bedeutsam wirken. Wir schauen also tatsächlich technisch gesehen in einen Spiegel: Was die Maschine antwortet, ist das, was ein Mensch antworten könnte. Die Fehler, die Lügen, die Faulheit – sie

¹ Vgl. «uncanny valley»-Effekt: https://de.wikipedia.org/wiki/Uncanny_Valley (10.5.2024)

² Vgl. dazu <https://www.ionos.de/digitalguide/online-marketing/web-analyse/turing-test/> (10.5.2024) Link zum Online-Turing-Test: <https://humanornot.so/> (10.5.2024)

³ <https://heycurio.com/> (10.5.2024)

⁴ Vgl. dazu: <https://schulesocialmedia.com/2012/11/21/rezension-jaron-lanier-gadget/> (10.5.2024) auch Laniers Webseite: <https://www.jaronlanier.com/> (10.5.2024)

sind alle über menschliches Verhalten in die Maschine reingeraten. Die Muster der Maschine sind die Muster der Menschen.

Die KI zeigt uns, wie wir selber funktionieren. Das ist wohl noch unheimlicher als ein Automat, der uns gleicht und den wir nicht von anderen Menschen unterscheiden können. Wir sind konfrontiert damit zu verstehen, was in uns eigentlich macht, dass wir denken und fühlen. Hat das etwas Maschinelles? Oder gar etwas Zufälliges?

Die KI-Forscherin Emily Bender spricht in Bezug auf Chatbots von stochastischen Papageien⁵, die einfach nach Wahrscheinlichkeiten Worte auswählen, die dann bedeutsam wirken. Die Nutzung von KI stellt uns also vor die Frage, ob wir nach bestimmten Mustern funktionieren, wenn wir sprechen, denken oder empfinden. Natürlich fühlt sich das für uns nicht so an, aber das, was die Chatbots hervorbringen, nehmen wir im ersten Moment auch nicht als etwas Maschinelles wahr. Wir sprechen mit ihnen, als wären sie ein menschliches Gegenüber. Siri, Alexa und der Stimme von GPS im Auto hören wir zu, als würden uns kluge Frauen durchs Leben führen. Und wenn wir ganz ehrlich sind: Beim Small Talk in der S-Bahn sagen wir zuweilen durchaus Dinge, die auch ein programmierter Papagei plappern könnte.

Hier kommt das Unheimliche ins Spiel: Wie ich mich bei einer Maschine täuschen kann, wenn ich denke, ich hätte es mit einem Menschen zu tun statt mit einem Automaten; so könnte ich mich auch in Bezug auf mein eigenes Denken, Sprechen und Handeln täuschen: Es könnte tatsächlich von einer Maschine hervorgebracht werden, ein Produkt von raffinierten Mechanismen und Mustern sein – aber nichts mit mir und dem, was sich in mir als Ich anfühlt, zu tun haben.

Ich erspare den Leserinnen und Lesern dieses Textes die billige Pointe, die man erzielen kann, indem man ChatGPT-Antworten kopiert und dann testet, ob alle rausfinden, was Mensch oder Maschine geschrieben haben. Aber ich nutze beim Schreiben eine Software, die nachverfolgt, welche Textpassagen ich reinkopiert habe und mich fragt, aus welcher Quelle sie stammen.⁶ Ich brauche also ein Hilfsmittel, um nachweisen zu können, dass ich einen Text geschrieben habe und nicht etwa eine Maschine. Bald wird diese Differenz verschwinden, wir werden annehmen, dass fast alle Texte automatisch generierte Textteile enthalten. Die Frage, wie sich dann mein Schreiben von dem von Maschinen unterscheidet, wird verschwinden. Dadurch wird immer weniger klar, was eigentlich menschlich ist und was nicht, wo die Differenz zwischen Robotern und denkenden Personen liegt. Ich schreibe dann als Teil eines riesigen Kollektivs und lasse von der Maschine Textmuster generieren, die andere Menschen so auch formulieren könnten. Nicht viel anders als die Kommentierenden, die mir mit Feedback geholfen haben, diesen Text schlüssiger und eleganter zu machen.

Wenn wir heute über die Möglichkeit maschineller Empathie, Moral oder Wahrheitsfindung nachdenken und durchaus auch etwas Angst empfinden, menschliche Errungenschaften könnten dadurch bedroht werden, wie wir KI einsetzen, dann können wir ausgehend von den bereits angestellten Überlegungen zwei Phasen unterscheiden und zwei Schlüsse ziehen: In einer ersten Phase ist KI ein Werkzeug, das wir einsetzen. Wenn dadurch Probleme entstehen, Unwahrheiten verbreitet werden, Fehler gemacht werden, dann ist all das Menschen zuzuschreiben, welche KI einsetzen. Würde ich mit diesem Text Lügen verbreiten, müsste ich dafür gerade stehen, auch wenn ich sie von ChatGPT übernommen hätte. Genauso wie ich sie verantworten müsste, hätte ich sie aus einem Buch abgeschrieben oder in einem Gespräch aufgeschnappt. Die Schlussfolgerung daraus wäre: Was Menschen mit KI machen, müssen Menschen rechtfertigen. Es hat mit Technologie nichts zu tun.

In einer zweiten Phase können wir aber Menschen und Technologie kaum noch unterscheiden. KI entwickelt Persönlichkeiten und wird schreiben und sprechen, wie Menschen das tun – während Menschen sich durch die intensive Nutzung von KI immer stärker wie Maschinen verhalten werden, die ja gleichzeitig auf menschliches Material zurückgreifen, um Musteranalysen durchzuführen.

Menschliches Verhalten wird dann mit maschinellem Tätigkeit so verbunden sein, dass alle Konzepte, von Wahrheit über Moral bis zu Liebe, sich einem Wandel unterziehen werden. Auch Verdrängtes und Tabuisiertes wird die Beziehung zwischen Mensch und Maschine beeinflussen. Heute wirkt das unheimlich, oft auch unerwünscht. Blickt man aber in die Geschichte der Menschheit, dann haben

⁵ <https://te.ma/art/6hqbo0/bender-gebru-sprachmodelle-stochastische-papageien/> (10.5.2024)

⁶ <https://app.originality.ai/home/dashboard> (10.5.2024)

sich alle Konzepte ständig verändert. Technologie und Innovation haben dazu geführt, dass wir so leben, wie wir heute leben. Und sie werden dazu führen, dass Menschen in Zukunft anders leben, als wir es heute tun. Der Spiegel der KI führt uns das aktuell gerade vor Augen. Das ist der zweite Schluss.

Die Angst vor einer KI, die autonom handlungsfähig wird und Menschen dabei bedroht, ist sehr verbreitet. Wir denken, getrieben von Science-Fiction-Serien, immer wieder über Systeme nach, welche die Kontrolle übernehmen und Menschen nicht mehr gehören. Diese populäre Angst ist aber lediglich eine Ablenkung von dem, was uns wirklich ängstigt: Dass Maschinen uns zeigen, wie wir Menschen funktionieren, wie wir (mit und ohne Maschinen) handeln – und wie wir uns verändern werden.