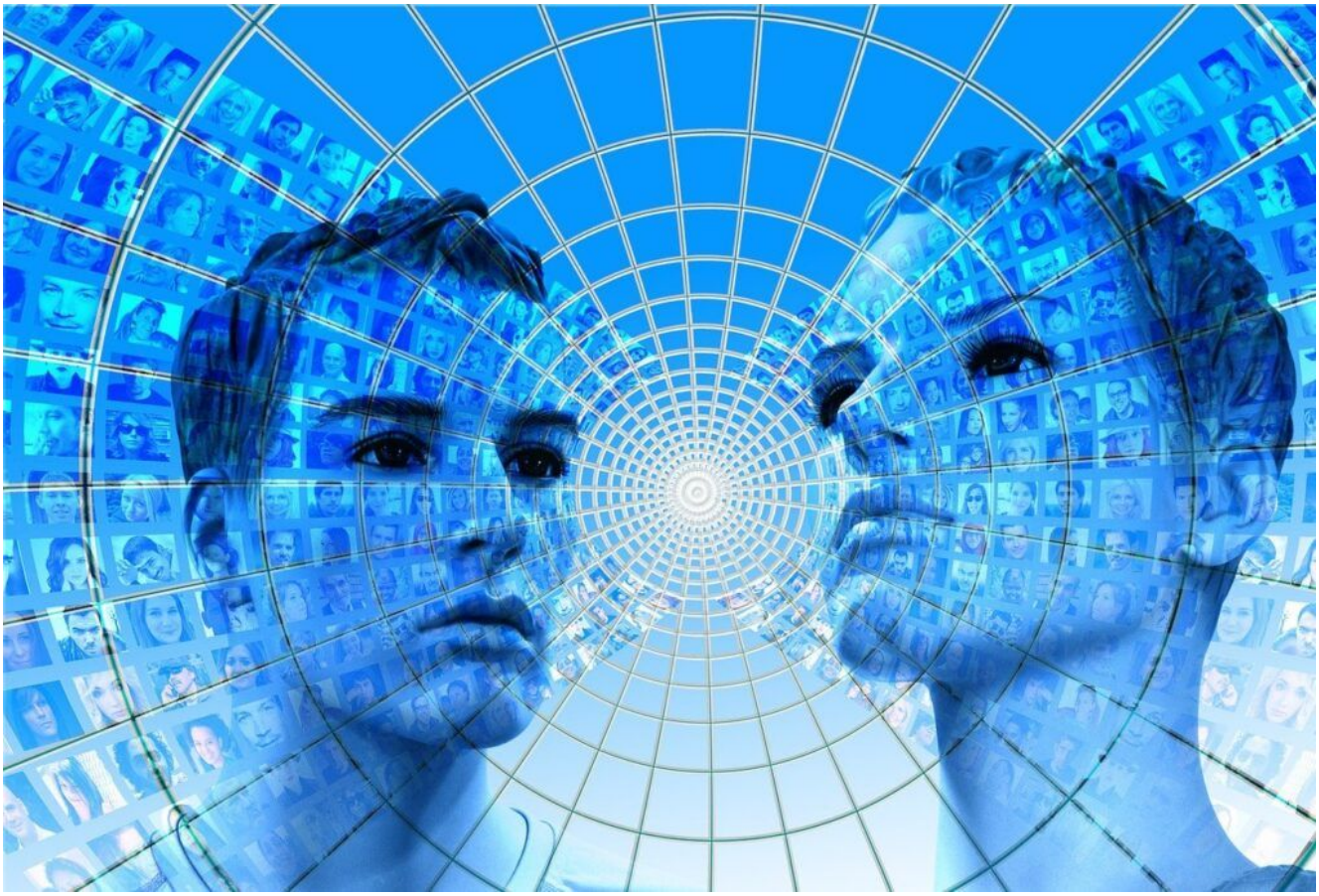
Seelen ex machina: Von der künstlichen Intelligenz zur künstlichen Psyche

Category: Blog
geschrieben von Gastautor | 22. Dezember 2023



Der Fall des Softwaredesigners Blake Lemoine ist noch in Erinnerung. Google feuerte ihn 2022, weil er glaubte, das Konversationsprogramm LaMDA («Language Model for Dialogue Application») spreche zu ihm. Lemoine twitterte: «Die Leute fragen mich immer wieder, warum ich denke, dass LaMDA empfindungsfähig ist. Es gibt keinen wissenschaftlichen Rahmen, in dem diese Feststellungen getroffen werden könnten (..) Meine Meinung über die Persönlichkeit und das Empfinden von LaMDA basiert auf meinen religiösen Überzeugungen».[1]



Edouard Kaeser ist Physiker, Philosoph, Jazzmusiker und freier Publizist.

Die Diagnose ist schnell gestellt: abgedrehter KI-Nerd. Aber interessanter ist Lemoine als Symptom einer Tendenz, die sich immer deutlicher im Umgang mit den neuesten KI-Systemen abzeichnet. Ich nenne sie «Umschalten auf eine Person».

Das Phänomen fiel in den 1960er Jahren dem Computerwissenschaftler Joseph Weizenbaum bei seinem Gesprächsprogramm ELIZA auf. Durch ein paar wirkungsvolle Tricks konnte es im Nutzer den Eindruck erwecken, es «verstehe» ihn und «gehe auf ihn ein». Weizenbaum war schockiert, wie leicht die Probanden am «personalen» Charakter des Programms festhielten, obwohl sie wussten, dass es sich um ein künstliches Gegenüber handelte.

Der Philosoph Ludwig Wittgensteins hatte das Problem schon früher in seinen «Philosophischen Untersuchungen» thematisiert: «Denke, ich sage von einem Freunde: ‚Er ist kein Automat'. – Was wird hier mitgeteilt, und für wen wäre es eine Mitteilung?» Unter normalen Umständen, so Wittgenstein, bedeutet dies, dass der Freund «sich immer wie ein Mensch, nicht manchmal wie eine Maschine benimmt». Und unter solchen Umständen ist «meine Einstellung zu ihm eine Einstellung zur Seele. Ich habe nicht die *Meinung*, dass er eine Seele hat».

Eine Seele haben – oder wie ich zu sagen vorziehe: eine Person sein – ist nicht eine Frage der Meinung, sondern des zwischenmenschlichen Zusammenlebens. Es mutet uns absurd an, von einem Menschen zuerst hypothetisch zu meinen, er

sei eine Person, und dann diese Hypothese zu testen, wenn möglich sogar zu verwerfen. KI-Systeme konfrontieren uns aber genau mit dieser Absurdität.

Im Film «Ex Machina» erhält der Softwardesigner Caleb vom Unternehmer und Milliardär Nathan den Auftrag, einen Turingtest mit der Roboterin Ava durchzuführen. Tatsächlich entspricht das Setting aber nicht jenem des Turing-Tests. Nathan umreisst den Auftrag so: Finde heraus, ob du immer noch das Gefühl hast, sie sei ein bewusstes Wesen, selbst wenn es sich um eine Maschine handelt. Caleb weiss, dass Ava ein Automat ist. Aber nach dem Umschalten hat der Automat den Status als «Seele» – als Person. Und «Zurückschalten» könnte sich unter Umständen als schwierig bis unmöglich herausstellen.

> *LaMDA begegnet uns als Konversationspartner, der Chatbot wird heute als ein künstlicher Autor gefeiert, DALL-E als künstlicher Illustrator.*



Kate Darling vom Media Lab des MIT (Boston): Ähnliche Entwicklung wie beim Umgang mit Tieren.

Werden wir also künftig immer mehr umschalten? Mit dieser Frage überschreiten wir die Schwelle zu einem neuen Zeitalter der Maschine. Wie Kate Darling vom Media Lab des MIT argumentiert, werden wir in den künftigen Mensch-Maschine-Interaktionen mit einem Kontinuum unterschiedlicher Einstellungen zu tun haben – ganz ähnlich wie in unserem Umgang mit Tieren, von den Amöben bis zu den

Bonobos.

Umschalten erscheint aus einem ganz bestimmten Grund notwendig. Genügend vielschichtige neuronale Netze sind Black Boxes. Sie haben Milliarden von Parametern, die sie automatisch einstellen. All die komplizierten Rechenprozesse vollständig computertechnisch zu beschreiben wäre ein praktisch unmögliches Unterfangen. Faute de mieux greifen wir zu psychologischen «Eselsbrücken», schalten um. Das ganze brachenübliche Marketing stellt ab auf vollmundiges Ankündigen von menschenähnlichen Fähigkeiten. Zum Beispiel «Watson kann alle Texte über Gesundheitsfürsorge in Sekunden *lesen*» (IBM), oder «Das KI-Modell hat den Menschen im *Verstehen* von natürlichen Sprachen überflügelt» (Microsoft).

Sprachliche Grossmodelle – wie der Generative Pretrained Transformer (GPT) – mögen stochastische Papageien und in diesem Sinn «seelenlos» sein, nichtsdestoweniger prägen sie unseren zwischenpersönlichen Verkehr. Ihre soziale Präsenz als interaktive Agenten ist schon nicht mehr wegzudenken. LaMDA begegnet uns als Konversationspartner, der Chatbot wird heute als ein künstlicher Autor gefeiert, DALL-E als künstlicher Illustrator. Der Nutzer dressiert dem KI-System seine Eigenheiten an, und rückwirkend imitiert das System den Nutzer auf eine Weise, die seinen Hang zum Umschalten verstärkt.

Sobald ein KI-System in unserem Netzalltag neue Normalität wird, gilt es nicht mehr als künstlich. Wir «bürgern» es «ein». Eine Kommission des Europäischen Parlaments schlug 2017 vor, gewissen Robotern den Status «elektronischer Personalität» zuzubilligen. Im gleichen Jahr verlieh Saudi-Arabien einer Roboterin das Bürgerrecht. Sollten wir also den Kreis von «unseresgleichen» erweitern und KI-Systeme als eine neue künstliche Spezies akzeptieren? Und könnte es dann sein, dass unser Empathieumfang auch künstliche «Seelen» einbegreifen müsste?

> *Ein solches Artefakt versteht uns nicht, es empfindet nichts, es sorgt sich nicht um uns, es simuliert einfach immer besser «Verständnis» für uns.*

Ob solchen Fragen mag einem blümerant zumute werden. Aber der Mensch hat gegenüber Geräten schon immer einen affektiven, ja, erotischen Hang erkennen lassen. Der Techno-Animismus ist ein altes Phänomen. Die Sozialpsychologin Sherry Turkle spricht von «Beziehungsartefakten». Ein solches Artefakt versteht uns nicht, es empfindet nichts, es sorgt sich nicht um uns, es simuliert einfach

immer besser «Verständnis» für uns. Das Problem, so Turkle, ist, dass wir uns gar nicht mehr daran stossen. Die kritische Feststellung «Da ist ja gar niemand zuhause» prallt ab an der Gegenfrage: «Ist denn bei dir jemand zuhause?»



Wir erblicken im Roboter vor uns den Roboter in uns, der Macht ausübt.

Beginnen wir dann auch unsere Seele als maschinell – als auf einer «Neuromaschine» laufend – zu begreifen? Ein Hirnforscher würde die Frage wahrscheinlich typisch ambivalent beantworten: «Nein», weil man viel zu wenig über den Zusammenhang von Gehirn und Psyche weiss; «im Prinzip ja», weil es durchaus erklärtes Ziel der Hirnforschung ist, diesem Zusammenhang auf den Grund zu gehen.

Wir beobachten das Umschalten ebenfalls in den zahlreichen Science-Fiction-Szenarien, mit ihren immer wieder variierten Schreckensvisionen einer Machtübernahme durch superintelligente Maschinen. Das Unterstellen solcher Motive spiegelt im Grunde das Unbehagen vor einem Teil unserer selbst. In unserem Unbewussten spielt sich vieles «maschinell» ab. Wir erblicken im Roboter vor uns den Roboter in uns, der Macht ausübt.

Die Romantiker ahnten dies alles voraus. In E.T.A Hoffmanns Erzählung «Der Sandmann» verdreht die Puppe Olimpia dem unglücklichen Jüngling Nathanael den Kopf, und sie stürzt ihn – nachdem ihre wahre künstliche Natur zutage getreten ist – in den Wahnsinn. Die modernen KI-Systeme treiben ihre Designer nicht in den Wahnsinn, sondern beflügeln ihre Phantasien über eine neue posthumane Spezies. Lassen wir dabei nicht ausser Acht, wer das grösste Interesse an der «Beseelung» der Maschine hat: die KI-Barone und ihre

Grossunternehmen. Die Künstliche-Seelen-Industrie hat bereits eine Investitionsexplosion ausgelöst. Und ihr Blick richtet sich auf die Kolonisierung unserer Seelen durch KI-Systeme.

Wohin dieses Abenteuer führt, ist ungewiss. Immerhin kann man aus der ganzen Entwicklung einen Appell heraushören: Lernen wir uns im Zeitalter einer posthumanen Spezies neu kennen! Gerade der Chatbot hat diesen Prozess jetzt angestossen. Er erweitert den Raum möglicher fremder Intelligenz und verschafft uns dadurch die Gelegenheit, unsere eigene Intelligenz, unser artspezifisches Mentalleben vergleichend neu zu erkunden: Was genau ist der menschliche Faktor? Das wäre das Projekt einer Anthropologie im Maschinozän. Wir geben bedenkenlos Milliarden für KI-Forschung aus. Wie viel investieren wir in jene Disziplinen, die sich wirklich mit Geist und Seele beschäftigen? Wann begreifen wir die Notwendigkeit eines komplementären human- und geisteswissenschaftlichen Programms, das den wahnhaften Wettlauf der KI-Systeme kritisch begleitet?

[1] https://twitter.com/cajundiscordian/status/1536503474308907010