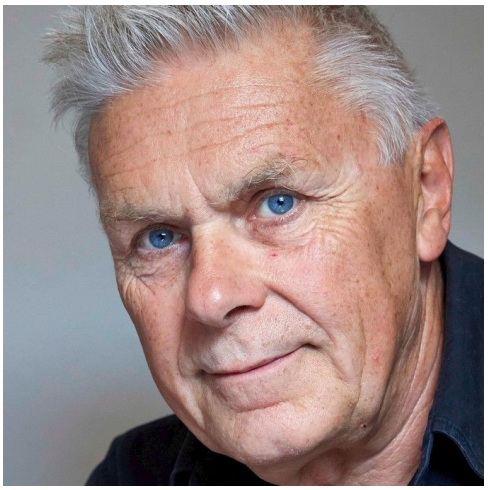
Falscher und echter Alarm in der KI

Category: Blog
geschrieben von Gastautor | 12. Dezember 2023

Das “Center for AI Safety” – eine amerikanische Nonprofit-Organisation – veröffentlichte vor kurzem ein Statement, das viele namhafte Fachleute unterzeichneten: “Das Risiko einer Auslöschung der Menschheit durch die KI sollte in der sozialen Dimension gleich eingeschätzt werden wie das Risiko der Auslöschung durch Pandemien oder durch Nuklearkriege”.[1]



Gastautor Eduard Kaeser

Wenn die KI wirklich eine solche Gefahr darstellt, fragt man sich, warum warnen dann ausgerechnet jene Leute vor ihr, die ihre Entwicklung rasend vorantreiben? Sind sie aus ihren Zauberlehrlingsträumen aufgewacht? Hat die "prometheische Scham" sie gepackt, das Gefühl der Unzulänglichkeit angesichts ihrer eigenen Produkte?

Natürlich sind die Warnungen nicht einfach in den Wind zu schlagen. Aber nüchtern betrachtet, eröffnen die KI-Systeme ein unerforschtes Feld der sozialen Interaktion von Mensch und Maschine, nunmehr auf "intelligentem" Niveau. Und es ist eigentlich nicht das KI-System selbst, sondern unsere vorauseilende dystopische Fantasie, die uns das Fürchten lehrt.

> *Wir machen uns a priori einen unzutreffenden Begriff von ihr. Und das führt zu Falschalarm.*

Karl Marx sprach vom "falschen Bewusstsein", also einem Begriffs- und Wahrnehmungsrahmen, der alles schon im Voraus präformiert, ja, verzerrt. Falsches Bewusstsein erscheint mir wie zugeschnitten auf die heutige KI-Technologie: Wir machen uns a priori einen unzutreffenden Begriff von ihr. Und das führt zu Falschalarm. Ich erläutere ihn kurz anhand dreier Punkte.

Anthropomorphe Analogien

Erstens attestieren wir den smarten Maschinen kognitive Fähigkeiten, deren Resultate wir oft nicht von menschlichen unterscheiden können. Wir sagen dann, der Chatbot "schreibe" einen Text, LaMDA (Language Model for Dialogue Application) "konversiere" mit uns, DALL-E 2 "male" ein Bild. Das sind aber nichts anderes als anthropomorphe Analogien, mit denen wir versuchen, die heute kaum noch vollständig durchschaubaren Maschinenabläufe in einem uns begreiflichen Idiom zu beschreiben. Insbesondere übertragen wir KI-Systemen bereits "moralische" Handlungsverantwortung, schieben ihnen etwa die Schuld zu, Jobs "abzuschaffen". Oder wir warnen vor "sexistischen" oder "rassistischen" Algorithmen. Als ob die Maschinen sich wie Menschen verhalten würden. Bisherige Klimax dieser KI-Beschwipstheit ist ein Softwaredesigner von Google – Blake Lemoine –, der behauptete, LaMDA sei in eine persönliche Beziehung zu ihm getreten.

Das Netz – eine Skinnerbox ohne Skinner

Hier zeigt sich – Punkt zwei – eine wirklich alarmierende Dialektik der ganzen Entwicklung. Im gleichen Zug, in dem wir Maschinen personenhafte Züge zuschreiben, vergessen wir die menschlichen Personen hinter den Maschinen. Nicht Maschinen schaffen die Jobs ab, sondern Unternehmen und Regierungen – Institutionen, die von Personen geführt werden. Und nicht die Algorithmen sind sexistisch oder rassistisch, vielmehr verwenden deren Designer und Benutzer sexistische respektive rassistische Daten.

> *Alarmierend an dieser Idiotie ist die naive Bereitschaft, unsere kognitiven, aber auch moralischen Kompetenzen an das smarte Gerät abzutreten.*

Der vermeintliche Machtwille des Robo sapiens kaschiert den Machtwillen des Homo sapiens. Man muss dabei nicht immer gleich China als abschreckendes Beispiel zitieren, wo das Regime die Technologie schamlos zur totalen Verhaltensdressur ausnutzt. Auch im Westen lassen wir uns von Algorithmen kontrollieren und manipulieren. Sie "entscheiden", was wir sehen, lesen, hören, kaufen, mit wem wir kommunizieren, wen wir mögen und wen wir hassen wollen. Das Netz ist eine gigantische Skinnerbox ohne Skinner. Wir selbst konditionieren uns zu Konsumratten.

Maschinen sozialisieren – Menschen maschinisieren

Das ist der dritte Punkt: Wir suchen die "intelligenten" Artefakte zu sozialisieren. Um Maschinen zu sozialisieren, muss man Menschen an sie adaptieren, "maschinisieren". Man achte einmal darauf, wie wir uns heute mit Apps armieren, die uns bei jeder Gelegenheit «beraten», was zu tun oder zu lassen sei. Eben lese ich die Werbung für die App "Blinklist" von Apple. Sie bringe "die Kernaussagen tausender Sachbücher auf das Smartphone. In nur 15 Minuten kann man sich so das Wissen eines dicken Sachbuchs aneignen".

Alarmierend an dieser Idiotie ist die naive Bereitschaft, unsere kognitiven, aber auch moralischen Kompetenzen an das smarte Gerät abzutreten. Das befördert eine Art von Techno-Fatalismus: Der Vormarsch der Algorithmen ist unaufhaltsam. Zurzeit warnen die Alarmglocken besonders schrill vor einer sogenannten künstlichen "Intelligenzexplosion". Die neuen Generationen von KI-Systemen könnten uns puncto Intelligenz überflügeln. Wobei sich der Verdacht hartnäckig hält, dass die Warner unter Intelligenz kaum mehr verstehen als ihre eigene nerdige Beschränktheit.

Das grösste Risiko: menschliche Unintelligenz

Künstliche Intelligenz ist eines der kühnsten, vielleicht das kühnste intellektuelle Abenteuer der letzten 60 Jahre. Nüchtern betrachtet, erweist sie sich als eine spezifische Form von Intelligenz, die ausschliesslich auf Rechenleistung – letztlich auf Energieverschleiss (Alarm auch hier!) – beruht. Sie kann uns viel darüber lehren, was ausser Rechenleistung sonst noch zur Intelligenz beiträgt.

Die seriöse Forschung zeigt uns zum Beispiel, dass KI ihre "intrinsischen" Grenzen hat. Und genau das ist spannend, Ansporn für weitere Forschung in der Entwicklung von KI-Systemen, etwa von biologienäheren neuronalen Netzen oder

von Quantencomputern. Zugleich aber auch dafür, die Alarmstufe herunterzufahren. Denn das grösste Risiko im Umgang mit der künstlichen Intelligenz bleibt die nicht durchschaute Unintelligenz des Menschen, vulgo: Selbstüberschätzung, selbst gratulierende Borniertheit, Inkompetenz-Erkennungs-Inkompetenz, Missions- und Machtwillen. Sie grassiert besonders ausgeprägt in KI-Kreisen.

[1] https://www.safe.ai/statement-on-ai-risk , Übersetzung E. K.