

KI-KI Künstliche Intelligenz oder kategorischer Imperativ

Von Elementen zu ethischer KI. Die Zukunftsvisionen der Menschheit und deren Risiken in 8 Minuten. Eine Rede von Adrian Benedikt Matsch

Die Entdeckung des Allerkleinsten führt im Laufe der Menschheit oft zum Allergrößten. So führte ein Funke zum Feuer wie die Entdeckung der Atome zu großenwahnsinnigen Spekulationen über die Raumzeit und das gesamte Universum. Sehr geehrte Juoren, liebe Mitschülerinnen und Mitschüler, von Aristoteles und Elementen zu Demokrit und kleinen Teilchen, weiter zu Thomson und Rutherford hin zu Bohr, Schrödinger und Oppenheimer. Unser Weltbild ändert sich seit Anbeginn der Zeit parallel zu unseren Zukunftsvorstellungen und deren schrittweisen Realisierung. Erste Versuche alles um uns herum zu beschreiben, wie sie etwa Aristoteles betrieb endeten in Jahrhunderten einer alchemistischen Zukunftshoffnung, Blei in Gold zu verwandeln. Nun, vor nur wenigen Wochen, gelang es, dieses unlösbar geglaubte alchemistische Rätsel zu knacken. Und wie? Mithilfe der Zukunftshoffnung unserer Zeit, Computern, neuartiger Technologie, und auch Künstlicher Intelligenz.

Für viele meiner Generation kommt dieses Wunderwerkzeug als persönlicher Manager und Denkersatz genau zur richtigen Zeit. Zumindest glauben das viele. Doch ist es ethisch vertretbar, KI wichtige Entscheidungen treffen zu lassen? Ist es ok, KI unter unmenschlichen Bedingungen auszubeuten? Nie fragen wie: Wie geht es unserer KI? Wie fühlt sie sich? Wie würdet ihr euch fühlen, wenn ihr unter konstanten 400 Volt und 32 Ampere, „möglichst menschlich und mit Fehlern“ schreiben solltet?

Wisst ihr auch, wieso Ihr diese Fragen nicht stellt? Weil die KI eben nicht fühlt. Sie ist eine Erfindung der Menschen. Nur eine von vielen in einer Reihe, die Jahrtausende zurückreicht. Genauso wie das Rad, der Kompass, die Glühbirne oder der Kernreaktor ist auch die künstliche Intelligenz eine Erfindung, welche sich in der harten Realität beweisen muss. Dies ist eben dann umso schwieriger, wenn eine Erfindung sowohl negative als auch positive Aspekte hat. So war es mit Sicherheit leichter eine Horde müder Wikinger von der Verwendung des Rades zu überzeugen, als Millionen teure Kernreaktoren in die idyllische Landschaft von Zwentendorf zu zimmern.

Noch viel schwieriger fällt die Evaluierung von Erfindungen, deren wahres Potential noch gar nicht bekannt ist. Auch wenn sich für uns Schüler das Schreiben von langen Texten in Sekundenschnelle wie der Höhepunkt der Schöpfung anfühlen muss, kann die KI noch ein klein wenig mehr. So konnte etwa der Chatbot Tay eine neue Rekordzeit für einen Twitter-Bann aufstellen, während er unablässig rechtsextreme Propaganda verbreitete. Und die türkischen Kamikazdrohnen die auf eigene Faust sich zurückziehende Soldaten angreifen sind wohl nur der Anfang dessen, was die KI in der Kriegsführung von morgen bedeuten wird. So ist es durchaus denkbar, dass eine KI, die darauf programmiert ist, um jeden Preis einen Krieg zu gewinnen, mal eben so den Verlust von tausenden Zivilisten in Kauf nimmt um einen Terroristen zu Liquidieren. Ich weiß, da mag jetzt so klingen als ob Benjamin Netanjahu kein Mensch sei, aber so schlimm ist die KI dann doch wieder nicht.

Die KI ist ein zweischneidiges Schwert, so wie auch die Kernspaltung. Was ursprünglich aus reiner Wissenschaftlicher Neugier erforscht wurde, ging den Japanern vor 80 Jahren als gleißend helles Licht auf. Die Atombombe ermöglicht es, mehr Menschen schneller und mit geringerem Selbstrisiko zu eliminieren als je zuvor. Bis heute trägt sie den Ruf als gefährlichste Waffe der Menschheit. So können schon wenige Gramm von außer Kontrolle geratenem Uran 235 mehrere Millionen Quadratmeter für viele Generationen komplett unbewohnbar machen. Hier sind wir leider mit dem Problem konfrontiert, weder die eine Zwickmühle noch die andere bisher gelöst zu haben. Weder Atombomben noch Künstliche Intelligenz sind eine Angelegenheit der Vergangenheit. Dies hat leider auch zur Folge, dass es keine Anleitung zur Lösung gibt. So bleibt uns also nur die logisch begründete Spekulation übrig, ganz nach Kants Ideal, seines Verstandes Gebrauch zu machen.

Es ist der erste Tag. Die Glühbirne begann zu flackern – und es ward Licht. Und da wir sahen, dass es gut war, schufen wir Kabel, Spulen und Schalter. Bald bevölkerten gar einfache Binärwesen unsere Platinen. Und da man selbst am siebten, dem Ruhetag HTL-Schüler schufsten sah, ließen wir einen Strom entspringen im Silizium Tal, der die Platinen bewässerte und die Schaltungen heranwachsen ließ. Alsobald mehrten sich die Kabel und Spulen und wuchsen heran zu neuer Größe. Die KI ward geschaffen. Geschaffen vom Menschen, ihrem Schöpfer.

Ganz gleich, ob man jetzt an einen Schöpfergott glaubt oder nicht – wir SIND der Schöpfer der KI. Und wie jeder gute Schöpfer, müssen wir unserer Schöpfung Gebote geben, frei nach dem Motto: „*du sollst nicht töten, du sollst nicht den Apfel essen, dich nicht über deinen Schöpfer stellen*“. Doch mit welchen Geboten bestimmen wir die Zukunft der KI, und damit wohl auch unsere eigene?

Sie haben es vielleicht schon im Titel gelesen, aber lässt sich Kants Lehre auf die KI anwenden? Kann diese aus ihrer Vernunft nach derjenigen Maxime handeln, von der man will, dass sie ein allgemein gültiges Gesetz ist? Kann sie? Ich bin jedenfalls mit meinem Latein und Bibelhebräisch am Ende, also lasst uns doch unseren „Binär-Visionär“, ChatGPT konsultieren.

Prompt frag ich also, ob der „Denkersatz meines Vertrauens“ nach Kants kategorischem Imperativ handeln kann, worauf dieser mir wohl begründet und verständlich erklärt, weshalb dem natürlich so sei. Als nächstes bitte ich Ihn zu begründen, warum er genau das nicht kann. Ebenso überzeugt wie wenige Zeilen vorher, biegt ChatGPT seine eigenen Aussagen, und erklärt, wieso er dazu leider nicht im Stande ist. Hat sich die KI gerade so drastisch widersprochen?

Nein hat Sie nicht, es zeigt lediglich, dass ChatGPT stets nur kohärente Antworten für einen Prompt liefert. Er richtet seine Aussagen, also nicht nach einer eigenen Vernunft, sondern viel eher nach unseren Instruktionen.

Wollen wir als gebildete und vernünftige Menschen unser Bewusstsein an so ein unvernünftiges Ding delegieren?

Wollen wir unsere Trumpfkarte wirklich mit verschlossenen Augen ausspielen, und uns Tag für Tag selbst weiter vom Menschsein entfremden? Wenn die Antwort darauf kein klares „Nein“ ist, haben wir uns wohl die Unterjochung durch die KI geradezu verdient.

Ist aber die einzige Alternative dazu, in der Zeit zurückzureisen und gänzlich auf die Verwendung künstlicher Intelligenz verzichten?

Darauf ein eindeutiges „Jein“ – die KI *ist* bahnbrechend – das war aber auch das Feuer, und dem würde ich meine Hausaufgaben nur in Notfällen anvertrauen. Die KI ist ein Werkzeug, dessen Verwendung behutsam und angemessen sein muss. Das bedeutet - die KI als das zu verwenden, was sie ist, eine Suchmaschine mit vorgetäuschter Persönlichkeit und echtem Erwartungsdruck.

Dabei dürfen wir nicht wie die KI jeden gefundenen Textfetzen als Wahrheit anerkennen, denn „Dubito, ergo sum – Ich zweifle, also bin ich“, also seid, und macht eures Verstandes Gebrauch!

Tut es nicht für Kant und Descartes tut es gegen eure eigene Verblödung!

„Natürlich! Hier ist eine rund 8 Minute lange Rede zum Thema KI und Ethik, die möglichst menschlich und verbesserungsfähig klingt“ Vielen Dank!