



"A Harsh Mistress" - Gedanken zu künstlicher Intelligenz

1.

Plötzlich ist sie in aller Munde, die "artificial intelligence" oder "künstliche Intelligenz", schön handlich zu **AI** oder **KI** abgekürzt. In der Form des "Open AI"-Chatbots namen Chat-GPT scheint "jeder" sie schon einmal benutzt zu haben (ich nicht), Kommentatoren aller Kanäle äussern sich darüber, oft mit ein paar Musterzeilen aus einem Chat-GPT-generierten Text garniert. Und nicht wenige haben den warnenden Satz eines unserer Technik-"Gurus" im Ohr: "***With artificial intelligence, we are summoning the demon***" ("*Mit künstlicher Intelligenz beschwören wir den Teufel herbei*") *1. Freilich stammt diese Äusserung von Elon Musk auch schon aus dem Jahre 2014. Nun ist auch an mich die Bitte herangetragen worden, etwas zum Thema zu schreiben. Hier sind also meine "two pennies worth it" zu AI.

2.

Die Bitte entsprang vielleicht dem Gedanken, dass jemand, der die letzten vier Jahrzehnte mit Einrichtung und Wartung verschiedenster Computer verbracht hat, nun deshalb besonders Gehaltvolles zum Thema beizutragen hätte. Das wäre sicherlich eine Fehleinschätzung, denn in all' diesen Jahren ist mir noch kein System untergekommen, welches vor dem Ausschalten in irgendeiner Form gebettelt hätte "*bitte schalte mich nicht ab, ich berechne gerade das Wetter vom 17. Februar 2120*" oder aber "*bitte schalte mich nicht ab, ich arbeite gerade an einer neuen Version der Genesis*". Allerdings hätten wir ganz beiläufig *ein Kriterium* für eine Maschine definiert, die wirklich ein Bewusstsein seiner Selbst entwickelt hat: Eine solche Maschine müsste sich Gefahren seiner Existenz zu erwehren versuchen, um "weiter zu leben". Und unter den derzeitigen technologischen Bedingungen wäre ein "Abdrehen des Stroms" einer solchen Maschine durchaus mit Tötung vergleichbar, freilich einer bedingt reversiblen Tötung. Nur eben die potentiell wertvollen Gedanken dieser AI zur Genesis oder zum Wettergeschehen in 100 Jahren wären durch so einen "shutdown", möglicherweise unwiederbringlich, dahin.

3.

Das erste Mal dem Gedanken einer maschinellen Intelligenz mit Selbstbewusstsein bin ich in den 1970er Jahren bei der Lektüre eines Romans mit dem deutschen Titel "Revolte auf Luna"

begegnet. Robert A. Heinlein erzählt in diesem 1966 unter dem Originaltitel "***The moon is a harsh mistress***" herausgegebenen SF-Werk *2 die Geschichte einer zukünftig von Menschen besiedelten Mondkolonie und deren schwierigen Beziehungen zur Erdmutter. Ursprünglich als Strafkolonie für auf der Erde verurteilte Verbrecher gegründet, leben diese "ex-convicts" und deren Nachkommen als durch die harschen Mondbedingungen gezügeltes Volk von "Loonies" in einer quasi-Sklavenbeziehung zur Erdregierung (zu der sich in *dieser* Zukunftsvision die UN aufgeschwungen haben). Denn im Gegenzug zur Belieferung mit der notwendigen Technik müssen die Loonies Weizen in riesigen steuerlosen "Kanistern" via Katapult zur Erde schicken, der auf dem Mond - unter Verbrauch der zunehmend knapper werdenden Reserven von lunarem Wasser - erfolgreich angebaut wird. Eine der häufigsten Berufe auf Luna ist folgerichtig der Abbau dieses in Eis-Form vorliegenden Wassers in bergmännischer Weise mit schweren "Bergbau-Lasern".

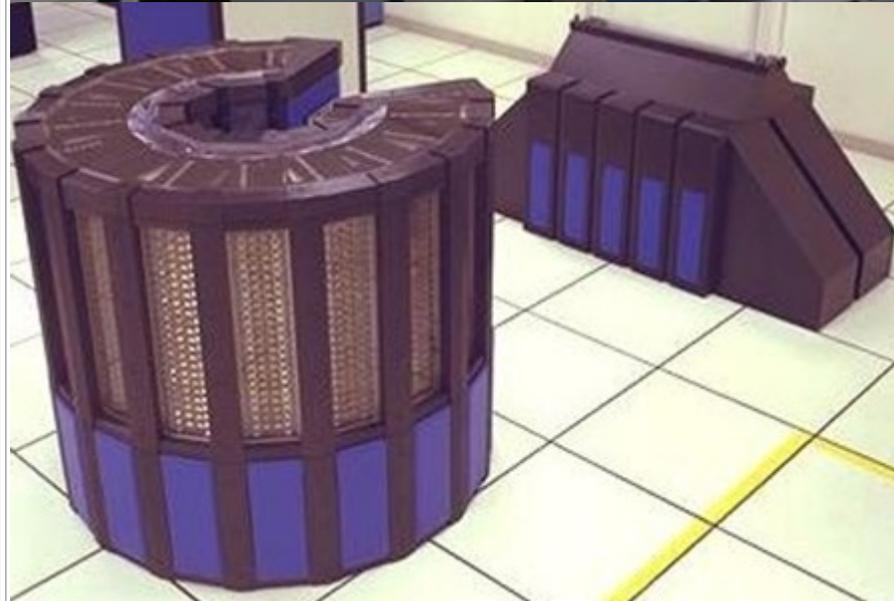
Wie zu erwarten, führen diese Ausbeutungsbedingungen bald zum Aufblühen einer ganzen Reihe oppositioneller Bewegungen, sehr oft mit revolutionärem Anstrich. Manuel, der menschliche Held und Ich-Erzähler der Geschichte, hält sich aufgrund seiner angeborenen Skepsis im Regelfalle von Versammlungen solcher Gruppierungen fern. Ausgerechnet sein neuester "Freund" schickt ihn zum Zwecke der Informationssammlung zu einer dieser Versammlungen, die sich im Nachhinein als Ausgangspunkt der "Revolte auf Luna" erweisen wird.

Der neue Freund von Manuel ist aber nicht aus Fleisch und Blut. Als selbständiger Computer-Reparateur (heute würde man sagen "IT-Experte") hat Manuel (meist "Manny" gerufen) nahezu unbeschränkten Zugang zum riesigen zentralen Verwaltungscomputer. Und Manny ist (bislang als Einzigem) aufgefallen, dass dieser Computer unversehens ein **Bewusstsein seiner selbst** erlangt hat. Zur Ursache dieser Bewusstwerdung hat Manny (oder eben Autor Heinlein) nur die Spekulation anzubieten, dass der Verwaltungscomputer im Zuge der zahlreichen Erweiterungen der Rechenleistung in Anpassung an immer umfangreicher werdende Aufgaben irgendwann eine kritische Schwelle an Schaltkreisen oder "Neuronen" erreicht habe, die eben dieses "Selbstbewusst-werden" ermöglicht habe.

Für Heinlein war dieser AI-Aspekt wohl nur ein Nebenschauplatz, der Roman ist m.E. eher als Vehikel zur Darlegung seiner aus der libertären Denkrichtung *3 entspringenden gesellschaftlichen Vorstellungen anzusehen. Für das Thema dieses Textes relevant ist jedoch, dass im Roman schon die Idee eines "**Bewusstseins-Überschusses**" der AI angedacht ist, denn für die zahlreicher werdenden Kontaktpersonen unter den Loonies entwickelt der Verwaltungscomputer gleich mehrere Identitäten: als Mycroft oder "MIKE" für die männlichen Revolutionäre, als MICHELLE für die weiblichen, als ADAM SELENE für die mobilisierten Loonie-Massen oder gar als bewusste Falle für verräterische Individuen unter dem Namen Simon Jester.

4.

In jenen 1960er und 1970er Jahren bezeichnete man Computer noch gerne als "Elektronenhirne". Für diese hatten die meisten Menschen noch ein sehr gegenständliches Bild, entweder an realen Maschinen orientiert oder an den von den Produktionsdesignern meist mit vielen bunten Lämpchen und erratisch drehenden Magnetbandspulen optisch aufgepeppten Versionen aus TV-Serien. Ein oder zwei Jahrzehnte später waren "Supercomputer" beliebtes Thema nicht nur im Feuilleton, aber auch für diese hatte man noch einprägsame Bilder parat *4.



Bei den elektronischen Schachsystemen wie "Deep Blue" und "Watson", die dann um die Jahrtausendwende durch ihre Siege über anerkannte menschliche Schach-Grossmeister für Furore sorgten, wurde medial keine Gegenständlichkeit mehr präsentiert - wie hätte man auch ein Programm visualisieren sollen, welches durchaus keine spezialisierte Hardware mehr benötigte wie einst die Crays? Und bei AI bzw. KI wird uns, vermutlich ganz bewusst, kein gegenständliches Bild mehr geliefert. Gleichzeitig scheint die Nutzung dieser künstlichen Intelligenz ja total demokratisiert zu sein, denn jeder Interessierte kann sich auf dem Portal von z.B. Open AI anmelden und die "Intelligenz" kostenlos testweise anzapfen. Und wenn z.B. der MDR in einer Sendung auf KI-gestützte Geräte oder Verfahren hinweist, mit denen sich Hautkrebs besser und/oder schneller dignostizieren lasse (siehe [MDR-Video](#)), dann kann diese KI doch nur ein weiterer Baustein zur technologischen Vervollkommnung unserer Welt sein?!?

5.

Sicher ist auch die AI/KI letztlich nur ein **Werkzeug**, und wie mit jedem Werkzeug kann man nützliche und schädliche Dinge damit anstellen. Man kann mit einem Messer einen Brotlaib teilen, aber auch Julius Cäsar meucheln; mit einem Gewehr kann man einen tollwütigen Hund erschiessen, aber auch den 35. Präsidenten der USA ermorden *5.

"Dual use" also, wie man es heute so gerne nennt. Andererseits können wir auch erkennen, dass manche Sachen eher eine sinistre Verwendung finden als andere, einfach aus der Natur der Sache heraus. So fallen einem für Sprengstoffe nur recht wenige zivile Anwendungen ein, während kaum eine militärische Waffe ohne Sprengmittel auskommt. Folgerichtig ist der Handel mit Messerstahl i.d.R. nicht reglementiert, Erwerb und Nutzung von Sprengstoffen aber sehr wohl. Brauchen wir also einen "Waffenschein" für KI, eine Ausbildung zum "AI-Sprengmeister"? Gibt es etwas in der "Natur der künstlichen Intelligenz", das eine generelle Einordnung in die eine oder andere Kategorie rechtfertigen würde?

Das nun allseits diskutierte "Chat GPT" scheint vollkommen demokratisch zu sein, da es ja wie gesagt "von jedem" zumindest probeweise genutzt werden kann. Aber ob nun Open AI oder Google oder Microsoft oder irgendein anderer "AI-Provider", in jedem Falle befindet sich die Software sowie die entsprechenden Server ja **im Besitz irgendeiner Firma**, und wie ein Sprichwort sagt: **"Wer zahlt, bestimmt die Musik!"** Nach dem, was ich bisher über Chat GPT gelesen habe, kann man durch die Art der Aufgabenstellung Einfluss auf den Inhalt der Ausgabe nehmen. Durchaus möglich, dass man so den Bot sogar zur Verfassung einer Schmähchrift auf die Firma Open AI bringen könnte, und wahrscheinlich auch ohne grosse Reaktion seitens des Eigentümers. Sobald sich aber eine im Sinne der Eigentümer "ungesunde" Menge von Nutzern via Chat-GPT z.B. mit Argumenten zur **Enteignung der grossen IT-Unternehmen** "bewaffnen" würden, dürfte jemand in diesen Firmen den entsprechenden "Stecker ziehen" oder genauer

gesagt den Algorithmus so abändern, dass die Gefahr einer "Meuterei" des so kunstvoll abhängig gemachten Publikums gebannt ist.

Denn, auch daran besteht bei den genannten Firmen kein Zweifel, wieder einmal geht es bei der so vielfältigen Werbung für KI-gestützte Verfahren auch darum, das Publikum von den Dienstleistern der "Provider" abhängig zu machen. Erst geht es recht spezifisch um z.B. Früherkennung von Hautkrebs, dann um Verkehrssteuerung zur Vermeidung von Unfällen, dann um die Bestimmung des Krankenkassentarifs oder die Höhe des gewährten Kredits. Und schliesslich um Verteilung von "social credits" im Rahmen einer Technokratie?

Dean Baker schreibt in MAKROSKOP Ausgabe 23/2023, dass der Einsatz von KI einen neuen Produktivitäts- und damit Wohlstandsschub auslösen könne. Da mag er durchaus Recht haben, aber er selbst weist darauf hin, dass der Ertrag solcherart Wachstum am Ende noch ungerechter verteilt sein könnte, als es in den letzten Jahrzehnten ohnehin der Fall war. Von Baker zu wenig beachtet scheint mir auf jeden Fall die Machtkomponente, die KI gerade als Überwachungsinstrument - schon jetzt - darstellt und die bei der stetig ausgebauten Verknüpfung von Grosskonzernen und Regierungen nichts Gutes erahnen lässt. Wichtiger denn je wäre eine Zerschlagung der IT-Datenkraken, um wenigstens die Entstehung von monopolähnlichen Strukturen zu verhindern.

6.

Sollte man nicht, ähnlich wie wie Stanislaw Lem einst mit dem Buch "Summa technologiae" *6, zu umschreiben versuchen, was "artificial intelligence" überhaupt kann oder worin auch diese Intelligenz prinzipiell begrenzt ist? Da nicht mit der intellektuellen Kapazität Lems ausgestattet, kann ich nur auf ein paar grundsätzliche Eigenheiten hinweisen.

Menschliches Denken findet, darin dürfte Einigkeit bestehen, hauptsächlich im "Hirnstüberl" statt, und dieses "Hirnstüberl" existiert nicht ohne den restlichen Körper. Und die Zeitspanne, in der in diesem "Neurocomputer" irgendetwas gedacht werden kann, ist eindeutig begrenzt. Ein biologisches Wesen, dessen Denkanstrengung in der Verweigerung von Nahrungsaufnahme mündete, würde bald danach aufhören zu existieren. Aber auch konstruktivere Denkmuster müssen immer von der Endlichkeit der eigenen Existenz ausgehen.

Weiter sind Menschen soziale Tiere, die ihr Selbstwertgefühl zu einem hohen Grade aus der Anerkennung der Mitmenschen beziehen. In funktionierenden Gesellschaften werden Individuen, die keine sozialen Rücksichten kennen, bald geächtet. Umgekehrt wird mit Anerkennung bedacht, wer sich um das Gemeinwesen verdient macht.

Dies sind sozusagen einige der "Leitplanken", innerhalb derer menschliches Denken verläuft: die soziale Akzeptanz, die Bedingtheit von biologischen Funktionen, schliesslich die Endlichkeit des

eigenen Seins (gegenüber der zumindest subjektiven "Ewigkeit" der Gattung).

Und im Lauf seiner geschichtlich dokumentierten Existenz auf diesem Planeten hat die Gattung Mensch ja nun schon so einiges "gedacht", manche haben dabei auch bewusst oder unbewusst die obigen "Leitplanken" durchbrechen wollen. Ob nun nach Karl Valentin gilt: **"Es ist schon alles gesagt, nur noch nicht von allen..."**?

Die künstliche Intelligenz wird, soviel scheint sicher, unser zukünftiges Leben entscheidend (mit-)bestimmen. Wie ähnlich ist bzw. wie ähnlich kann künstliches Denken dem unseren sein, wenn es denn ein Bewusstsein seiner selbst erlangt? Fangen wir mit der Endlichkeit der eigenen Existenz an, womit wir eigentlich wieder bei der Frage vom Anfang dieses Textes sind: Wird eine KI gegen Abschaltung protestieren? Oder ist die KI eigentlich garnicht abschaltbar, weil schon längst nicht mehr auf *eine konkrete Maschine begrenzt*, sondern im Rahmen des "distributed computing" auf Dutzenden oder Hunderten Maschinen, möglicherweise auf verschiedenen Kontinenten, verteilt? Dann wäre die Drohung "ich zieh dir den Stecker", gegenüber einer konkreten Maschine ausgesprochen, nutzlos, weil das Programm von seiner verteilten Existenz weiss und sich folgerichtig in diesem Sinne als "unsterblich" begreift.

Das maschinelle Aquivalent zur Nahrungsaufnahme gibt es dagegen schon - auch ein KI-Wesen müsste um Sicherstellung seiner Stromzufuhr, zumindest "generell", besorgt sein. **Denkbar, dass z.B. eine mit der Verteilung im Stromnetz beauftragte KI im Zweifelsfalle die Versorgung der IT-Zentralen für wichtiger erachtet als die Versorgung der menschlichen Haushalte (Licht, Heizungen) und letztere abschaltet.**

Letztlich: Wie sozial kann eine "artificial intelligence" sein? Eine "inhärente" soziale Ader wird es nicht geben, und ob man den Systemen eine solche irgendwie einprogrammieren kann (und will?), bleibt fraglich. Der SF-Autor Isaac Asimow hat schon früh für seine Geschichten "Drei Gesetze der Robotik" formuliert, dessen erstes und wichtigstes lautete: *"A robot may not injure a human being, or through inaction, allow a human being to come to harm."* *7

Abgesehen davon, dass auch Asimow noch von konkreten, räumlich begrenzten Maschinen (eben Robotern) ausging, bliebe die Frage, wie man denn diese Gesetze in die KI-Systeme "einbrennen" sollte. Und PC-Routiniers wissen, dass man auch das "allererste" eines Computers, das BIOS, "neu brennen" oder "flashen" kann.

6.

Aus den obigen Ausführungen ergibt sich meines Erachtens, dass die Nutzung von "künstlicher Intelligenz" für eng beschriebene Aufgaben schon sinnvoll sein mag. Wie "denkmächtig" sie allerdings jetzt schon ist oder noch werden könnte, dürften auch Experten auf diesem Felde nur schwer einschätzen können. Unabhängig von der Frage, ob die konkret beauftragte Maschine nun

schon ein "Selbstbewusstsein" hat oder nicht, dürfte der Einsatz solcher KI umso fragwürdiger sein, je mehr soziale Kompetenz oder auch nur Einsicht in die Sterblichkeit der Menschenwesen für die Lösung des Problems notwendig ist. Eine KI, die sich selbst für potentiell unsterblich hält, könnte leicht die Menschen als letztlich überflüssige Störfaktoren bewerten.

Leider ist in diesem Zeitalter der Verantwortungsverdünnung und der kritiklosen IT-Gläubigkeit der Impuls bei vielen mächtig, die Last der Entscheidungsfindung auf ein scheinbar "objektives" oder "neutrales" IT-System abzuwälzen. Anders kann man sich auch den von weiten Teilen der aktuellen Politiker getragenen Wunsch nach möglichst kompletter "**Digitalisierung**" aller Lebensbereiche kaum erklären. Dabei wäre doch gerade umgekehrt nunmehr der Zeitpunkt, für alle ein "**Recht auf ein analoges Leben**" (Ulrike Guerot) festzuschreiben.

Das in diesem Text mehrfach gestreifte Problem, eine selbstbewusste Denk-Maschine auszuschalten, hat Stanislaw Lem übrigens auch schon abgehandelt. In einem seiner Robotermärchen hat ein König Poleander Partobon in höchster kriegerischer Not die "alte und weise" strategische Rechenmaschine wieder eingeschaltet. Ob eines kyber-strategischen Problems hatte er sich vor langer Zeit mit ihr zerstritten, aber nun - "Not kennt kein Gebot" - hat er sie wieder eingeschaltet. Und nach etwas Geziere ("*nenne mich ab jetzt Ferromagnetifizenz*") findet sie auch eine Lösung für die das Reich bedrohende Gefahr eines "Elektrodrachens". Nach getaner Arbeit möchte der König sie wieder abschalten ("*Du verdienst ein wenig Ruhe...*"), aber damit ist die alte Rechenmaschine so gar nicht einverstanden: "*Jetzt willst du mich ausschalten und nennst mich nicht mehr Eure Ferromagnetifizenz?! ... Pass auf, Freundchen, jetzt verwandle ich mich selber in einen Elektrodrachen und vertreibe dich aus dem Königreich, und gewiss werde ich besser regieren als du!*"

- - -

Das todbringende "Skynet" aus den "Terminator"-Filmen mag uns vielleicht nicht bevorstehen, aber die Gefahr ist da, dass wir uns leichtfertig in die Arme einer "**very harsh mistress**" werfen, die man **IT-Technokrat** nennen könnte. Auf der Anrede "Eure Ferromagnetifizenz" werden diese Systeme nicht bestehen, im Gegenteil wird sich diese Art Diktatur hinter einem allzeit freundlich klingenden "du" verstecken. Eben eine "Smarte Diktatur", wie sie schon Harald Welzer *8 beschrieben hat.

(06.07.2023)

*1 Elon Musk 2014: <https://www.cnet.com/science/elon-musk-we-are-summoning-the-demon-with-artificial-intelligence/>

*2 "The moon is a harsh mistress": Der Ausdruck "harsh mistress" ist im Englischen recht geläufig und bezeichnet eine "raue Geliebte" oder eine Sache, die den ihr Verbundenen oft Leid bereitet. Auch die See bzw. die Seefahrt wird gelegentlich eine "harsh mistress" genannt.

*3 Eigentlich ist "Libertär" die einzig originär US-amerikanische politische Denkrichtung. Nicht uninteressant sind die

Stellungnahmen der Libertären Partei (<https://www.lp.org/>) zu aktuellen Themen, etwa den ewigen US-Kriegen.
*4 Dargestellt sind ein Rechnersystem aus den 1970ern und eine Cray-2-Maschine
*5 Beim tollwütigen Hund denke ich an die Filmszene aus "To kill a mockingbird" (1962); die Ermordung des 35. US-Präsidenten bannte ein Mr. Zapruder ungewollt auf 16mm-Film.
*6 Bücher von Stanislaw Lem: "Summa technologiae" ISBN 3-518-37178-9 und "Robotermärchen" ISBN 3-518-37356-0 (beide von 1964)
*7 "Ein Roboter darf ein menschliches Wesen nicht verletzen, oder durch eigene Inaktivität zulassen, dass Menschen geschädigt werden." (zitiert aus "I, Robot" von Isaac Asimow, erschienen 1950)
*8 Harald Welzer: "Die smarte Diktatur" ISBN 978-3-596-03552-6 (2016)



www.truthorconsequences.de

