

Bullshit ist ein Feature, kein Bug



„Ein Beispiel dafür wäre etwas, was in der Biologie fast immer hilfreich ist: mehr Energie zu erhalten. „Das erste, was passieren könnte, ist also, dass ein solches System sagt: ‚Wir brauchen mehr Energie. Lasst uns den ganzen Strom zu meinen Prozessoren umleiten.‘ Ein weiteres großes Unterziel wäre dann, mehr Kopien von sich selbst zu machen. Hört sich das für Sie gut an?“ (aus William Hertling: [Singularity](#) – via Heise-Forum)



Ich darf das roboteraffine Publikum auf einen [Heise-Artikel](#) über [Geoffrey Hinton](#) aufmerksam machen (leider Paywall). „Warnt vor Gefahren“ ist natürlich langweilig, das hat schon Stanislaw Lem getan, vor allem in [Ananke](#), vor mehr als einem halben Jahrhundert. Die „Gefahr“ ist eher, so Lem, dass die Geschöpfe der Robotik uns ähnlicher sind bzw. sein werden als zu wünschen wäre. (Was sagt KI zum Ukraine-Krieg?)



Geoffrey Hinton, credits: [playgroundai.com](#)

Hintons Sicht der Dinge wurde maßgeblich von der neuen Generation großer Sprachmodelle verändert, insbesondere GPT-4 von OpenAI, das im März heraus kam. Es habe ihm klar gemacht, dass Maschinen auf dem Weg sind, viel schlauer zu werden, als

er dachte, sagt er. Es beunruhigt ihn, wie sich das entwickeln könnte. „Diese Dinger sind völlig anders als wir“, sagt er.“

Das wage ich aus philosophischer Sicht zu bezweifeln. Der Mensch hätte, wenn KI – in welcher Form auch immer – so anders wäre als er selbst, unbewusst etwas geschaffen, dass er dann auch nicht verstehen könnte. (Darüber muss ich noch nachdenken. Was [sagt Hegel?](#))

Das Ziel sind [selbst lernende neuronale Netze](#). Die sind aber nicht anders als das menschliche Gehirn, nur ausgelagert, [wie jedes andere Werkzeug auch](#).



„Unsere Gehirne haben 100 Billionen Verbindungen“, sagt Hinton. „Große Sprachmodelle haben bis zu einer halben Billion, höchstens eine Billion.“ Doch GPT-4 wisse Hunderte Male mehr als jeder Mensch.“

„Wissen“ ist aber nur ein technisches Problem. Hätte ein Mensch in einer Nanosekunde alles im Internet vorhandene Wissen zur Hand, wäre er genau so schlau. Das Problem ist doch eher, wie man damit umgeht und wie man es einordnet.

Hal, stelle mir alle verfügbaren Quellen zusammen, die den Übergang von der Sklavenhaltergesellschaft zum Feudalismus ökonomisch erklären, [insbesondere die Spezifik](#), warum ein Zusammenhang bestehen könnte zur Herausbildung des spezifische

Kapitalismus in Nordwesteuropa. Oder: Hal, gibt es Klassenkampf im [□□□□□□□□](#)?



Bullshitting sei ein Feature, kein Bug. „Menschen konfabulieren immer“, sagt er. Halbwahrheiten und falsch erinnerte Details seien Kennzeichen der menschlichen Konversation: „Konfabulation ist ein Merkmal des menschlichen Gedächtnisses.“ Diese Modelle machten damit, sagt Hinton, etwas genauso wie Menschen. Der Unterschied bestehe darin, dass Menschen normalerweise mehr oder weniger korrekt konfabulieren. Das Erfinden sei nicht das Problem. Computer brauchen einfach ein bisschen mehr Übung.



Meine zwei Cents dazu: KI wird den Kapitalismus revolutionieren wie schon das Internet, ihn aber nicht

abschaffen. Roboter sind Teil der Produktivkräfte, nicht mehr. Aber vielleicht gibt es ja doch eine Überraschung, wenn der erste Roboter anfängt, Karl Marx zu lesen...