

## *Wittgenstein (LW) und Künstliche Intelligenz (KI)*

Motto:

„Es ist schon ein großer und nötiger Beweis der Klugheit oder Einsicht, zu wissen, was man vernünftiger Weise fragen solle. Denn, wenn die Frage an sich ungereimt ist, und unnötige Antworten verlangt, so hat sie, außer der Beschämung dessen, der sie aufwirft, bisweilen noch den Nachteil, den unbehutsamen Anhörer derselben zu ungereimten Antworten zu verleiten, und den belachenswerten Anblick zu geben, dass einer (wie die Alten sagten<sup>1</sup>) den Bock melkt, der andre ein Sieb unterhält.“

Kant: *Kritik der reinen Vernunft* B 82-3.

Wie jedermann, der meine website mit philosophischen Texten ([www.emilange.de](http://www.emilange.de)) auch nur angesehen hat, wissen kann, habe ich mich ausführlich für LW interessiert und aus seinen Konzeptionen eine eigene philosophische Konzeption entwickelt: reflexive Klärung der formalen Begriffe des Alltagsverstehens<sup>2</sup>. Nicht interessiert habe ich mich bisher für KI. So wie LW die Mengenlehre als eine Kombination von Beweismethoden und Begriffsverwirrung und die empirische Psychologie für eine Kombination von experimentellen Methoden und Begriffsverwirrung gehalten hat, so hielt ich KI für eine Kombination von informationstechnischen Konstruktionen und Begriffsverwirrung.

Die sich mit KI beschäftigende Seite des Feuilletons meiner morgendlichen *Süddeutschen Zeitung*<sup>3</sup> ist geeignet, das für die folgende Erörterung zu ändern und über LW & KI zu handeln.

### I.

Zuerst eine kurze Beschreibung der Seite. Sie enthält drei auf mein Thema bezügliche Einheiten. In den Spalten 3-5 oben eine von einem Computer-Programm (>Dall-E<) erstellte, Comic-artige Zeichnung zum Thema >Run-DMC und Wittgenstein machen einen Turing-Test<.<sup>4</sup> In den Spalten 1-2 über die ganze Seite einen Artikel unter dem Titel >Gute

---

1 Polybios: *Annalen* XXXIII, 21, 1.

2 Vgl. >Philosophie<; >Philosophy< auf der website.

3 7. Dezember 2022, Seite 9.

4 >Run-DMC< war eine amerikanische hip hop Gruppe auf Queens, Ney York. Kommentar: Sie bestand aus drei Mitgliedern, die Zeichnung zeigt nur zwei. >Turing-Test' war ein von dem englischen Mathematiker Alkan Turing vorgeschlagenes Prüfverfahren, bei dem der Proband vor einem undurchsichtigen Vorhang sitzen sollte, hinter dem

Unterhaltung< von Andrian Kreye. In den Spalten 3-6 bis zum letzten Drittel der Seite ein Interview mit einem KI-System namens GDP<sup>5</sup> zu seiner Funktionsweise und seinen Beschränkungen.

Der Artikel von Kreye hat den Untertitel „Man hat streckenweise den Eindruck, man chatte mit einem menschenähnlichen Wesen: Was die künstliche Intelligenz hinter Chat GDP draufhat“. Sowohl er als auch sein Kollege Bovermann zeigen die gebotene Skepsis gegenüber der „Menschenähnlichkeit“ des Chat-Bots. Der Automat >sagt< selbst: „Ich existiere lediglich als eine Sammlung von Algorithmen und Daten, die es mir ermöglichen, Text zu verarbeiten und zu generieren.“ GDP soll sich von Vorgänger-Modellen nur durch die gesteigerte Qualität des Outputs unterscheiden.

Dennoch zeigen sie sich von den angeblich unabsehbaren Konsequenzen der neuen Stufe von Textverarbeitung und -generierung beeindruckt (Bovermann mehr als Kreye) und glauben, es werden sich Frage wie die nach der >Bewusstseins<-Fähigkeit und etwaigen Personenrechten stellen (es gibt freilich auch genug Philosophen, die das glauben).

Ich möchte argumentieren, dass die Autoren mit ihrer Skepsis ganz Recht haben und sogar noch viel weiter gehen müssten. Selbst „Menschenähnlichkeit“ wäre nicht genug für „den Moment der Singularität ... , an dem eine künstliche die menschliche Intelligenz erreicht“ (Kreye). In kürzester Weise begründet: Das liegt daran, dass es die Natur des Menschen ist, nicht nur Mensch, sondern Person zu sein. Mit dem Personen-Begriff, auf dessen exklusive Unterscheidung vom Begriff eines Gegenstandes unser normales Begriffs-System aufgebaut ist [es gelten die Regeln: >Wer jemand=eine Person ist, ist nicht etwas (in dem auf logisch Einzelnes bezüglichen Sinn).< Und: > Was etwas ist, ist nicht jemand.<] Mit dem Personen-Begriff aber kommt notwendig (a priori) ein Paket von Begriffen, das mit ihm intern verknüpft ist und das auf Computer nicht angewendet werden kann und über das sie nicht verfügen *können*.

Die Autoren haben also Recht mit ihrer Skepsis, aber mobilisieren für sie viel zu schwache Gründe. Das liegt daran, dass sie sich in ihren (eigenen) Begriffe nicht gut genug auskennen. Dem kann Philosophie abhelfen und damit der im Motto angeführten, von Kant gesehenen Gefahr der Dummheit = des Mangels an Urteilskraft abhelfen.

---

eine Turing-Maschine (ein sehr einfacher Computer) und eine Person sitzen sollten. Wenn der Proband zwischen den Äußerungen der beiden Entitäten hinter dem Vorhang nicht unterscheiden könnte, sollte das ein >Beweis< dafür sein, dass die Turing-Maschine – ja, was – Bewusstsein hat, eine Person ist ...etc.? Kommentar: Die Zeichnung zeigt keinen Vorhang. Dass der Turing-Test >inconclusive< bleiben muss, hat in einem einfachen Gedanken-Experiment (>The Chinese room experiment<) der Sprachphilosoph John Searle gezeigt.

5 Kurz für >Generative Pre-trained Transformer<.

## II.

Was hat das Argumentationsvorhaben mit LW zu tun? Zweierlei: Er ist die Quelle einiger Einsichten, deren ich mich für meine Argumentation bedienen werde. Und die Autoren, die ich kritisiere, gehen selbst auf LW ein.

Kreye berichtet davon, dass man GDP die Aufgaben gestellt habe, einen Text über LW im Stile verschiedener Popsänger zu verfassen.<sup>6</sup> Dies seien die Ergebnisse gewesen: „Born in Austria, Raised on the streets of Vienna, Wittgenstein was a philosopher, With a Mind always a-trying. He tackled the big questions, Of language and truth and reality.“ (*Bruce Springsteen*) „Wittgenstein on the mic, Rapping 'bout language and thought. He's the king of philosophy. And he can't be bought.“ (*Run-DMC*) „His words are like a hammer, slamming down on the brink. Let us raise our fists. And scream out his name, Wittgenstein, Wittgenstein, The philosopher of flame.“ (*Metallica*) „Wittgenstein, oh Wittgenstein, Your words are like a burning fire.“ (*Bob Marley*)

Zumindest der ersten beiden >Texte< im Stil von *Springsteen* und *Run-DMC* sind nicht nur expressiv und sagt etwas Richtiges über LW, in NES-philosophischer Form (wie NES-Kaffee mit viel heißem Wasser zu etwas sehr Kaffee-Ähnlichem verwandelt werden kann, so könnten diese Texte mit viel Erläuterung und Einbettung zu einer richtigen Auskunft über LW gemacht werden).<sup>7</sup>

## III.

Auch wenn es in Goethes >Faust< Mephistopheles ist, der dem Schüler >Zuerst Collegium Logicum< rät, und er dann von der Logik wenig schmeichelhaft redet<sup>8</sup>, ist der Rat jedem nützlich, der sich in seinen eigenen Begriffen besser auszukennen lernen will. Ich werde daher eine wichtige und nicht geläufige logische Unterscheidung LWs an den Anfang stellen und entschuldige mich für den lehrhaften Ton. Aber wo etwas erst zu lernen ist, ist lehrhafter Ton

---

6 Mir hat das auch deshalb Spaß gemacht, weil mein eigener größter Erfolg die popkulturelle Verbreitung meiner Entdeckung zur formal-ästhetischen Struktur der *Logisch-Philosophischen Abhandlung* (LWs einzigem zu seinen Lebzeiten veröffentlichten Buchs) in einem Wissenschafts-Comic des Psychiaters John Heaton und der Zeichnerin Judy Groves gewesen ist (>Wittgenstein for Beginners, ICON BOOKS 1994, 31 mit Nachweis 175).

7 Die Rede von >NES-Philosophie' stammt von Wolfgang Fritz Haug, einem Marx-Forscher an der FU Berlin zu Zeiten meines Studiums, bei dem ich ein Descartes-Seminar besucht habe. Später war seine Marx-Lektüre ein Thema meiner kritischen Untersuchungen zu Marx (>Das Prinzip Arbeit<, Berlin 1980).

8 *Der Tragödie Erster Teil*, Vers 1911 ff.

nicht leicht zu vermeiden.

Hier ist eine logische Einsicht LWs, die ich mir zunutze mache: *Materiale* oder inhaltliche Begriffe, mit denen etwas in Wahrnehmung oder tätigem Umgang Gegebenes charakterisiert werden kann, müssen von *formalen* Begriffen unterschieden werden. Wörter, die in bestimmten Verwendungen formale Begriffe ausdrücken, sehen nur so aus wie Begriffswörter, die materiale Begriffe ausdrücken. Tatsächlich drücken sie Variablen aus, Begriffe höherer Stufen, die anderwärts schon Klassifiziertes (begrifflich Charakterisiertes) zusammenfassen.<sup>9</sup>

An dieser Stelle bediene ich mich der Einsicht eines zweiten großen Philosophen des 20. Jahrhunderts, W.V.O. Quine.<sup>10</sup> Er hat darauf hingewiesen, dass die Variablen in formalen Notationen (wie der variable Name  $\langle x \rangle$  in der Prädikatenlogik 1. Stufe) in der Umgangssprache Vorformen haben – die indefiniten Pronomina. Der formalen Variable für Gegenstände,  $\langle x \rangle$ , entspricht beispielsweise das indefinite Pronomen  $\langle \text{etwas} \rangle$ .

Über LW und Quine hinaus geht nun meine These, dass die Umgangssprache ein zu  $\langle \text{etwas} \rangle$  gleichrangiges indefinites Pronomen (die Vorform einer Variablen) enthält, nämlich  $\langle \text{jemand} \rangle$ . Wie  $\langle \text{etwas} \rangle$  dem formalen Begriff  $\langle \text{Gegenstand} \rangle$  entspricht, so  $\langle \text{jemand} \rangle$  dem formalen Begriff der  $\langle \text{Person} \rangle$ .

Damit ist zugleich meine Behauptung begründet, dass unser alltägliches Begriffs-System über der Dualität der formalen Begriffe von  $\langle \text{Gegenstand} \rangle$  und  $\langle \text{Person} \rangle$  aufgebaut ist. Das Argument ist: Wofür wir indefinite Pronomina [oder (Vorformen von) Variablen] ausgebildet haben, mit dessen distinkter Existenz rechnen wir jedenfalls.<sup>11</sup>

#### IV.

Ich habe bisher nur eines von zwei Merkmalen des Begriffs  $\langle \text{formaler Begriff} \rangle$  erwähnt, die LW in der *Logisch-Philosophischen Abhandlung* erwähnt: dass er eine Variable darstellt (4.126 h). Das andere ist: dass er mit jeder seiner Einsetzungsinstanzen (mit jedem Wert der Variablen) schon gegeben ist (4.12721). Es ist leicht einzusehen, dass formale Begriffe, weil sie zuvor schon Klassifiziertes zusammenfassen, mit dem Klassifizierten auch schon (zumindest implizit) gegeben sind. Wenn wir Wörter für natürliche Objekte (Steine, Äpfel etc.) und für Artefakte (Tisch, Fortepiano, Computer) haben, bräuchten wir den Begriff eines

<sup>9</sup> Vgl. zum Begriff eines formalen Begriffs im Ganzen: *Logisch-Philosophische Abhandlung* 4.122-4.128.

<sup>10</sup>  $\langle \text{On what there is} \rangle$  in: *From a Logical Point of View*, 1-19.

<sup>11</sup> Die gewundene Formulierung ist dem Umstand geschuldet, dass Quine's Grundsatz „To be is to be the value of a variable.“ ( $\langle \text{On what there is} \rangle$ , l.c., 15) als allgemeine Erklärung für singuläre Existenz umstritten geblieben ist.

Gegenstandes nicht explizit, weil er spätestens mit der Bedeutungserklärung solcher Wörter ins Spiel käme (>Ein Apfel ist ein *Gegenstand*, der an einem Baum bestimmter Sorte wächst und bei Reife gepflückt werden oder abfallen kann.<)

Eine Folgerung aus dem Status formaler Begriffe und ihrem zweiten Merkmal ist, dass formale Begriffe nicht leer sein können. Es ist keine Situation denkbar, in der nach der Existenz von etwas gefragt werden könnte und in der es nicht Gegenstände und Personen gäbe (Personen auch, denn wer fragt oder antwortet, muss eine Person sein). Das ist bei materialen Begriffen anders. Der Ausdruck >geflügeltes Pferd< mag biologisch unmöglich sein, logisch widersprüchlich ist er es nicht – aber er ist immer leer (d.h. es finden sich keine Anwendungsexemplare für ihn).

## V.

Nun ist, bevor dieses Kapitel Collegium Logicum geschlossen werden kann, noch eine Komplikation zu vergegenwärtigen.

Sowohl >Gegenstand< als auch >Person< drücken (in bestimmten Verwendungen) formale Begriffe aus, deren Werte im ersten Fall die Bezeichnungen von natürlichen Objekten oder Artefakten sind; im zweiten Fall in erster Linie die Namen von Personen [in zweiter Linie Indikatoren (>ich/du; er/sie; wir/ihr<) und in dritter Linie Kennzeichnungen (>der/die so-und-so<), die auf Personen zutreffen und mit denen auf sie Bezug genommen werden kann]. Aber sie unterscheiden sich voneinander darin, dass >Person< auch einen materialen Begriff ausdrückt, >Gegenstand< in korrekter Verwendung niemals. (Der Film-Titel lautete nicht >Zwei oder drei Gegenstände, die ich von ihr weiß<, sondern >Zwei oder drei Dinge (frz. choses) ...<.)

Als einen materialen Begriff ausdrückend bedeutet >Person< >ein sprechendes, tätig/handelndes und sich selbst bewertendes Lebewesen<. Hier wäre nun wieder für jedes Merkmal eine begriffliche Begründung erforderlich. Die will ich aber hier nicht geben, sondern verweise auf meine Untersuchungen zum Personen-Begriff.<sup>12</sup>

---

<sup>12</sup> Im entsprechenden Kapitel von >Philosophie</>Philosophy<; in >The concept of a person and the ontology of common sense<; sowie in: >Philosophie des Personalbewusstseins – Personidentität im Alltagsverständnis<. Alles auf [www.emilange.de](http://www.emilange.de).

## VI.

Es ist Zeit, sich dem Thema KI und dem möglichen Personen-Status oder der Bewusstseinsfähigkeit von Fortentwicklungen von GDP und vergleichbaren Konstruktionen zuzuwenden.

Der Ausdruck >künstlich< weist auf den Artefakt-Charakter der entsprechenden Konstruktionen hin. >Intelligenz< heißt, wie immer die Humanwissenschaften versuchen mögen, den ausgedrückten Begriff zu operationalisieren und damit zu präzisieren, lexikalisch einfach >Einsichtsvermögen<. Das ist sicher eine komplexe und sektoral gegliederte Fähigkeit oder ein Bündel von Fähigkeiten *von Personen*. Die können auf einem Gebiet Einsichtsvermögen haben (z.B. in begriffliche Zusammenhänge, wie ich), auf anderem kann es ihnen weitgehend fehlen (z.B. auf dem der höheren Mathematik, wie ebenfalls mir).

Es gibt mit dieser Erklärung bereits zwei begriffliche Gründe, warum Computer logisch keine Personen sein können: Sie sind keine Lebewesen, sondern von Lebewesen hergestellte Artefakte. Und sie haben kein Einsichtsvermögen. Denn man wird >Einsicht> sicher von bloßem Wortverstehen unterscheiden wollen, etwa durch den Zugang zu den Gründen für eine in einem Satz ausgedrückte Meinung/Auffassung. Gründe und Ursachen sind die einfachsten sachlichen Kontexte für Sätze/Meinungen/Auffassungen und für GDP sollen „Empathie .., Kontext, kritisches und ethisches Denken“ z.B. unzugänglich sein.

Bleibt als zweiter Kandidat für die Charakterisierung von KI nach dem Erreichen des Moments der Singularität die Bewusstseinsfähigkeit. GDP äußert sich dazu folgendermaßen:

„Ich bin ein Sprachmodell und habe kein Bewusstsein im Sinne von Selbstbewusstsein oder Empfindungen. Ich bin nur ein Computerprogramm, das dazu entwickelt wurde, auch natürliche Sprache zu antworten.

.....

Bewusstsein ist keine Frage der Datenverarbeitung. Die Fähigkeit zu Bewusstsein ist ein komplexes und umstrittenes Thema in der Philosophie und Wissenschaft.“

In einer hier nicht ausführlich zitierten Passage >glaubt< GDP aber, „dass künstliche Intelligenzen immer besser werden und möglicherweise in der Lage sind, komplexere Formen von Bewusstsein zu entwickeln“.

Das fordert folgende kritischen Kommentare heraus. Wie >Einsichtsvermögen< ist >Bewusstsein< gewiss ein komplexes Konzept. Bemerkenswerter Weise gibt es nur zwei alltägliche Redensarten, in denen der Ausdruck, der ja ursprünglich ein philosophischer Kunstausdruck war, eine ungezwungene Verwendung hat: >bei Bewusstsein sein< bzw. es (nach einer Ohnmacht etwa) wiederzuerlangen. Da ist klar, wie komplex das Konzept ist. Denn wieder bei Bewusstsein sein heißt, wieder ansprechbar sein, wieder seiner Glieder mächtig zu sein, wieder jede seiner erworbenen Fähigkeiten betätigen können etc. Die zweite Sorte von Verwendungsweisen sind die einer propositionalen Ergänzung bedürftigen intentionalen: >er/sie ist/war sich bewusst, dass....<. Ich glaube, dass in beinahe allen Fällen dieser Verwendungen der Ausdruck >bewusst sein< zugunsten des Ausdrucks >wissen< eliminiert werden kann. Und für >wissen< genügt in diesen Kontexten die schon platonische Erklärung als >begründete wahre Meinung<.

Außerdem besteht Bedenken gegen die Äußerungen der KI GDP, die mit >ich< anfangen. Der Indikator steht für die 1. Person und wird legitim nur von natürlichen Personen nicht-eliminierbar gebraucht. Man hat da nur die Möglichkeit, entweder zu leugnen, dass GDP wirklich *ich* sagt, oder das es *ich sagt*. Das erste, weil die wirkliche Verwendung von >ich< das Bewusstsein seiner Ersetzbarkeit durch den Personennamen des Äußerers und auf ihn zutreffende Kennzeichnungen voraussetzt. Das zweite, weil >sagen< hier >zu verstehen geben/behaupten< bedeutet und das intentionale Tätigkeiten sind, die GDP mit dem sich selbst >absprechen< von Selbstbewusstsein ausschließt. Es ist Selbstbewusstsein, dass sich im >ich<-Sagen äußert.

Man könnte sich freilich trotz aller dieser Bedenken einfach (begrifflich) *entscheiden*, einer KI nach dem Erreichen des Moments der Singularität Bewusstsein zuzuerkennen. Denn schließlich fassen auch andere formale Begriffe oft sehr Heterogenes zusammen.

Beklagenswerter Weise geht philosophische Argumentation oft so aus – wie das Hornberger Schießen. In Ermangelung von genauerer Information über Hornberg und ein Schießen dort muss man sinnvoller Weise annehmen, dass dieses unentschieden ausgegangen ist.

© E.M. Lange 2022