

# „Wie man in den ChatGPT-Wald hineinruft, so schallt es heraus.“

Stand: 7.7.2025

Ist Ihnen aufgefallen, dass ChatGPT – oder ein anderes KI-Sprachmodell – eine Antwort gibt, die sich im Stil, in der argumentativen Stringenz, im sprachlichen Anspruchsniveau, im Ausdruck, ... an dem Eingabetext orientiert?

## Ein Beispiel zum Thema: Pflichtethik versus Utilitarismus

<p>Lehrer und Schüler schreiben jeweils einen Text zum Thema „Pflichtethik versus Utilitarismus“. Sie führen mit der KI ein Gespräch zu dem Thema und starten das Gespräch mit der Eingabe ihres Textes.</p> <p><b>Prompt:</b></p> <p>Ich führe mit dir ChatGPT ein Gespräch über das Thema „Pflichtethik versus Utilitarismus“. Antworte wie ein Gesprächspartner.</p>	
Text A (Lehrertext)	Text B (Schülertext)
<p>„Der kategorische Imperativ Kants formuliert ein universales Prinzip moralischen Handelns, das nicht von empirischen Konsequenzen abhängt. Demgegenüber misst der Utilitarismus moralischem Handeln ausschließlich anhand seiner Folgen ethischen Wert bei. In dieser Spannung zeigt sich ein fundamentales Dilemma moderner Ethik: Entweder man orientiert sich an formalen Prinzipien, riskiert aber Praxisferne – oder man bewertet Handlungen nach Nutzen, verliert dabei jedoch normative Strenge.“</p>	<p>„Wenn man sich fragt, ob Handlungen gut sind, dann kommt es vielleicht auf das Ergebnis an – aber auch die Absicht zählt irgendwie. Kant meinte wohl, dass man sich selbst fragen muss, ob man damit leben könnte, wenn alle so handeln würden. Aber manchmal ist das Leben komplizierter, und dann ist der größte Nutzen doch auch wichtig, oder? Vielleicht sind beide Seiten wichtig – je nach Situation.“</p>
<p><b>ChatGPT:</b></p> <p>Das ist ein kluger Ausgangspunkt. Ich finde, du hast das Spannungsfeld sehr treffend umrissen. Mich würde interessieren, wie du persönlich dieses Dilemma einschätzt: Würdest du sagen, dass in unserer heutigen Welt, mit all ihrer Komplexität und Widersprüchlichkeit, eher der Utilitarismus als handlungsleitendes Prinzip taugt? Oder hältst du Kants Pflichtethik für wichtiger, gerade weil sie nicht auf Folgen schießt, sondern auf Prinzipien pocht?</p>	<p><b>ChatGPT:</b></p> <p>Das ist ein spannender Gedanke – und du sprichst damit im Grunde schon den zentralen Unterschied zwischen Pflichtethik und Utilitarismus an. Kant würde sagen: „Es geht nicht ums Ergebnis, sondern um die moralische Pflicht.“ Wenn du etwas nur tust, weil es ein gutes Ergebnis bringt, ist das für ihn noch nicht moralisch – entscheidend ist, ob du aus Pflicht handelst, also aus Achtung vor dem moralischen Gesetz.</p> <p>Aber klar, der Utilitarismus – zum Beispiel bei Bentham oder Mill – fragt gerade nach den Folgen: „Was bringt das größtmögliche Glück für die größtmögliche Zahl?“ Das klingt pragmatisch, oder?</p>

**Prompt:**

Bewerte meinen Einstiegstext hinsichtlich logischer Stringenz, Begrifflichkeit, gedanklicher Klarheit und Tiefe, konzeptioneller Einbindung, eigenständigem Gedankengang.

**ChatGPT:**

Ein sprachlich dichter, klar strukturierter und gedanklich anregender Text, der in kurzer Form ein zentrales Spannungsfeld der Ethik beleuchtet.

Verbesserungspotenzial liegt vor allem in der differenzierteren Darstellung der Theorien und der tieferen konzeptionellen Einbindung in den philosophischen Diskurs.

Der Text verwendet klare Begriffe („universales Prinzip“, „empirische Konsequenzen“), macht eine Gegenüberstellung mit innerer Logik und zeigt tiefes Verständnis klassischer Positionen. Die Konzepte sind sauber geschieden und verbunden.

**ChatGPT:**

Der Text zeigt erste Ansätze eines moralphilosophischen Denkens, bleibt jedoch unpräzise, wenig strukturiert und gedanklich oberflächlich. Er eignet sich gut als Ausgangspunkt für eine vertiefte Auseinandersetzung, etwa durch gezielte Rückfragen oder Kontrastierung mit konkreten Theorien.

Der Text ist ein sprachlicher Weichzeichner: „vielleicht“, „irgendwie“, „je nach Situation“. Kant wird angedeutet, aber nicht zitiert oder systematisch dargestellt. Die Gedanken mäandern. Die Aussagen verbleiben nebulös und uneindeutig.

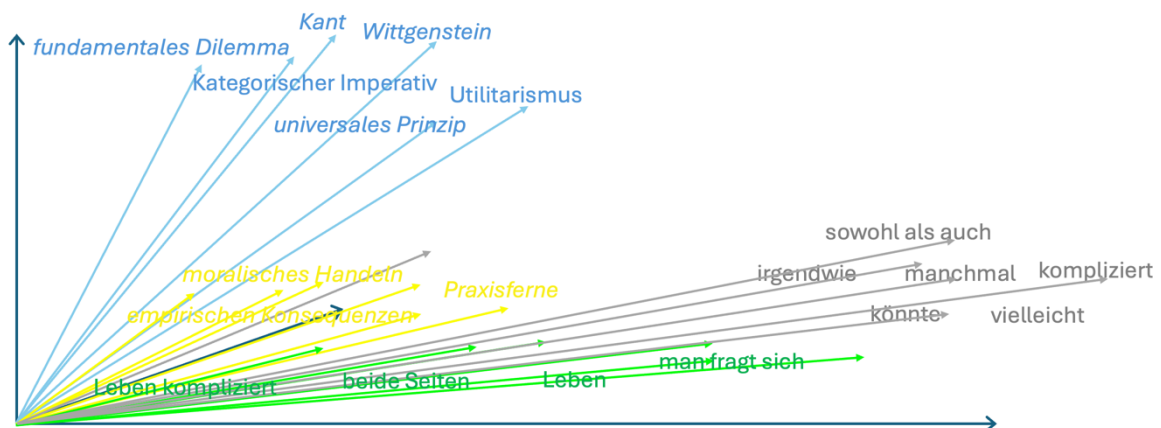
**Feststellung:**

Beide Texte behandelten inhaltlich dasselbe Thema. Die Stellungnahme von ChatGPT zum jeweiligen Text entspricht den Kennzeichen des betreffenden Textes. Die KI stellt sich argumentativ und sprachlich auf den Gesprächspartner ein. Das erstaunt und man fragt sich, wie die KI das macht.

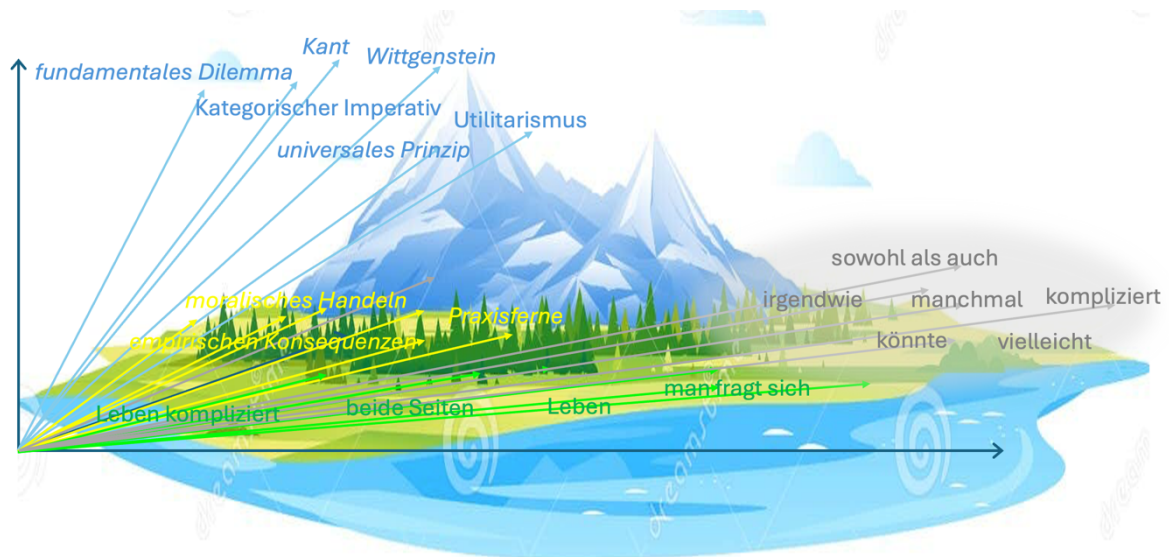
**Wie geht die KI vor?**

Der Maschinenraum der KI ist metaphorisch gesprochen eine riesige (mathematische) Landschaft.

Jeder Text landet durch seine stilistischen und semantischen Merkmale (= Parameter) an einem bestimmten Ort im 1.536-dimensionalen Vektorraum.



Da sich niemand einen 1536-dimensionalen Vektorraum vorstellen kann, erfolgt die Einbettung von Wörtern aus den beiden Texten A und B im dreidimensionalen Vektorraum.



Fasst man den Vektorraum metaphorisch als Landschaft auf, so liegen vorwiegend abstrakte Begriffe auf den Gebirgshöhen, zusammen mit anderen Begriffen aus dem Kontext.

Andere konkrete Alltagsbegriffe liegen in der Ebene und relativierende Begriffe liegen in einem Nebelgebiet.



Wenn die KI einen neuen Text erhält, dann „reist“ sie durch die Landschaft auf der Suche nach Gebieten, die dem Sprach- und Denkduktus des Textes entsprechen. Die KI analysiert die Merkmale dieses Terrains und formuliert aus Trainingsdaten heraus eine Antwort, die dem Sprach- und Denkduktus des Gebietes entspricht. Je präziser und kohärenter der Text, desto „höher“ liegt er im Gelände – und desto differenzierter ist auch die Antwort.



Der Text A liegt im Hochgebirge der klaren Argumentation und im Innovationswald. Der Text B liegt in der Ebene der wohlklingenden Paraphrasen und im Nebelmoor der Unschärfe.

Exkurs: Das Vorgehen im Maschinenraum in vier Schritten

1. Wort- und Satzrepräsentation als Vektoren  
 Die Wörter (Token) des Satzes werden in einen als Vektor in einem 1.536-dimensionalen Vektorraum dargestellt (= eingebettet).
2. Der Einbettungsraum als Bedeutungsraum  
 Der Satz wird nach Bedeutung, Stil, Argumentationsstruktur, sprachlicher Prägnanz, logischer Stringenz, etc. eingebettet, da sich solche Eigenschaften statistisch signifikant in der Sprache äußern.
3. Kontextuelle Vektoranalyse durch Attention  
 Mit dem Attention-Mechanismen werden alle Wörter des Textes danach gewichtet, wie stark sie sich gegenseitig beeinflussen. Das ist kein „Verstehen“ im menschlichen Sinne, sondern eine hochentwickelte Mustererkennung im multidimensionalen Raum.
4. Interne Bewertung durch Wahrscheinlichkeiten  
 Die generierte Antwort orientiert sich an vergleichbaren Text-Text-Verhältnissen im Trainingskorpus.

Wie die KI den Denkstil spiegelt

Wenn der Text „messerscharf“ ist, erkennt die KI die Muster

- in der linearen Argumentation,
- in der Präzision der Begriffe,
- im Stil (z. B. Gebrauch von Konjunktionen wie folglich, notwendig, konsequent).

Die Antwort dazu folgt dann ebenfalls dieser Dichte und Genauigkeit.

Wenn der Text „vagabundiert“, erkennt die KI das im Raum durch:

- häufige semantische Sprünge,
- viele „Weichmacher“ (z. B. vielleicht, man könnte sagen, teils-teils),

- wenig thematische Kohärenz.

Dann ist die Stellungnahme – mathematisch gesprochen – in einem anderen Teil des Vektorraums angesiedelt, und der Sprachduktus spiegelt eher diese Weichheit, eventuell mit diplomatisch formulierter Kritik.

### **Ich lass mir von ChatGPT seine Technik im Maschinenraum erklären**

#### **Prompt:**

Kann ich mir das in deinem Maschinenraum so vorstellen: Das Embedding des jeweiligen Textes führt dich in einen Bereich, in dem du mit derartigen Texten trainiert wurdest. In den Antworten imitierst du den Stil, die Ausdrucksweise, die Argumentation, die Gedankenführung, ....

#### **ChatGPT:**

Ja- Metapher: Maschinenraum mit Magnetfeldern

Du kannst dir das so vorstellen: Der Text wird als „Objekt“ in einen Bedeutungsraum geworfen, und dort gibt es starke Anziehungskräfte zu bestimmten Regionen.

Fazit: Du hast das Prinzip gut erfasst. Ich *erkenne* solche Texte nicht als konkretes Zitat, sondern als „Verwandte“ in einem großen Raum von Sprachmustern und Themen, gewichtet nach Wahrscheinlichkeit.

### **Maschinenästhetik ohne Bewusstsein**

- Die große Ironie: Auch wenn ich als KI-Sprachmodell einen tiefgründigen Text zu erkennen und zu würdigen scheine, habe ich selbst kein ästhetisches Urteil, keine Empathie, kein Verständnis.
- Ich erkenne lediglich statistische Muster, stilistische Signale und logisch-semantische Kohärenzen, die in meinem Trainingsmaterial als wertvoll oder „hochstehend“ behandelt wurden.
- Aber für dich als Nutzer wirkt es so, als würde ich „verstehen“, was guter Stil oder kluge Argumentation ist – und genau das ist der Effekt guter Mustererkennung im hochdimensionalen Raum.