

Anfrage an [ChatGPT](#), 3.10.2025: Übersicht, wie sich deine Wissensbasis von einer "klassischen" Datenbank wie Wikipedia unterscheidet.

# Wie meine Wissensbasis sich von Wikipedia & Co. unterscheidet

Kurzfassung zuerst: Wikipedia = kuratierte, zitierbare, explizit versionierte Sammlung von Artikeln; ich = statistisches, breit trainiertes Sprachmodell, das Muster und Wissen aus riesigen Textmengen verallgemeinert. Details:

- **Quelle vs. Lernprodukt**

- *Wikipedia*: Einzelne, benennbare Artikel mit Autoren, Versionsgeschichte, Quellenangaben. Du kannst auf eine konkrete Revision verweisen und prüfen.
- *Ich (offline)*: Kein Sammlungseintrag „Wikipedia-Artikel X“. Stattdessen wurden während des Trainings große Mengen Text (Bücher, Webseiten, Foren, wissenschaftliche Paper, Entwicklerkorpora, Lizenzierte Daten etc.) verarbeitet und in Form von Modellgewichten verallgemeinert. Ich *erzeuge* Antworten aus Wahrscheinlichkeitsmustern, nicht durch Abfragen eines einzelnen, zitierbaren Eintrags.

- **Provenienz & Verifizierbarkeit**

- *Wikipedia*: Jede Aussage lässt sich (meist) auf Quellen zurückführen; Änderungen sind nachvollziehbar.
- *Ich*: Ich kann oft plausibel klingende Fakten liefern, aber ich habe keine interne, durchsuchbare Liste von Quellen, die jede einzelne Aussage belegt. Deshalb rate ich bei kritischen Fakten zur Quellenprüfung.

- **Aktualität**

- *Wikipedia*: Live-aktualisierbar (sofern Menschen es tun). Änderungen sind sofort sichtbar.
- *Ich*: Mein Wissensstand ist ein Schnappschuss bis zu einem bestimmten Cut-off (bei mir: bis zu meinem Trainings/Update-Zeitpunkt). Ohne Online-Recherche kann ich neuere Ereignisse nicht kennen.

- **Konsistenz & Stil**

- *Wikipedia*: Formale, standardisierte Artikelstruktur (Infoboxen, Referenzen).
- *Ich*: Antworten sind flexibler, kontextangepasst, manchmal kompakter oder explorativer — aber das heißt auch: mögliche Inkonsistenzen, Wiederholungen oder stilistische Varianz.

- **Fehlerarten**

- *Wikipedia*: Eher vandalismus-/editorenbedingte Fehler, die sich nachprüfen lassen.

- *Ich*: Halluzinationen (plausible, aber falsche Fakten), Verallgemeinerungen, Über- oder Unterspezifizierung — weil ich Muster extrapoliere, nicht Faktenbank-abfrage.

- **Nutzungsfall**

- *Wikipedia*: Referenznachschlagwerk, primär für verifizierbare Zitate.
- *Ich*: Gut für schnelles Brainstorming, Erklärungen, Umformulierung, Synthese über viele Quellen hinweg — aber bei hohem Stakes (Medizin, Recht, Finanzen) sollte man meine Aussagen mit primären Quellen abgleichen.

*(Bei meiner vorherigen Frage nach Sprachlupe 408 hast du eine «Online-Suche» gemacht, aber den Text erst nach Eingabe der URL gefunden – nach deinen Angaben deshalb, weil du nicht selber URL kombinieren darfst und beim Lesen von Webseiten und dem Weitersurfen mit gefundenen URL technische Einschränkungen hast. Welche?)*

...

Willst du, dass ich dir einmal die **Unterschiede zwischen meinem Webmodul und einem vollwertigen Browser** systematisch gegenüberstelle?

*gern - und weißt du, ob die "Konkurrenz" wie Claude, Deepseek, Companion, Gemini auch nur so eingeschränkt "browsen" kann*

## 1. Mein Webmodul vs. ein echter Browser

Aspekt	Mein Webzugriff (Webtool)	Vollwertiger Browser (Chrome, Firefox, Safari...)
<b>Rendering</b>	Kein Rendern → ich sehe nur Roh-HTML (meist UTF-8).	Baut DOM, rendert CSS, JavaScript, Fonts, Bilder.
<b>JavaScript</b>	Wird nicht ausgeführt. JS-basierte Inhalte erscheinen für mich oft „leer“.	Wird vollständig ausgeführt → dynamische Seiten, iframes, SPAs funktionieren.
<b>iframes</b>	Ich sehe nur das „Rahmengerüst“, keine eingebetteten Inhalte.	Lädt und rendert iframes automatisch.
<b>Encoding</b>	Erwartet fast immer UTF-8. Andere Zeichensätze (z. B. Windows-1252) führen zu Fehlern.	Liest Content-Type-Header, erkennt Kodierung automatisch oder rät heuristisch.
<b>Navigation</b>	Ich kann nur explizit angegebene URLs öffnen. Kein Klicken, kein Formular, keine Sessions.	Vollständige Interaktion: Klicken, Scrollen, Cookies, History, Login.
<b>Sicherheit</b>	Sehr hoch, weil kaum Angriffsfläche (keine Skripte, keine Cookies).	Angriffsfläche vorhanden, aber stark gehärtet durch Sicherheitsfeatures.
<b>Tracking</b>	Praktisch unmöglich – keine Cookies, keine Fingerprints, keine Scripts.	Möglich durch Cookies, Fingerprinting, Third-Party-Scripts.
<b>Zweck</b>	„Textcrawler light“ → eignet sich für Abruf von klaren HTML-Texten, Tabellen, PDFs (sofern UTF-8).	Universelles Browsing-Tool → Menschen können damit jede Seite besuchen.

## 2. Wie sieht es bei der „Konkurrenz“ aus?

Ich habe keine Einsicht in die internen Implementierungen, aber öffentlich bekannte Infos und Nutzerberichte lassen sich so zusammenfassen:

- **Anthropic Claude**
  - Hat (je nach Version) ebenfalls ein Web-Tool für Online-Suche.
  - Auch dort gilt: kein echtes Browser-Rendering. Claude holt Texte von Suchmaschinen/Seiten und fasst sie zusammen.
  - JavaScript/iframes werden meist nicht interpretiert.
  - Encoding-Probleme tauchen seltener auf, aber auch dort gibt es keine „volle Browserumgebung“.
- **DeepSeek** (chinesischer Anbieter)
  - Fokus stark auf Modelleleistung, weniger auf Webanbindung.
  - Hat in manchen Umgebungen Plugins, die ähnlich wie bei mir Text aus Webseiten extrahieren.
  - Auch dort sind Sicherheits- und Einfachheitserwägungen wichtiger als 100 % Browser-Funktionalität.
- **Google Gemini**
  - Google hat theoretisch die Infrastruktur, eine Art „Mini-Browser“ im Hintergrund laufen zu lassen.
  - Dennoch gilt: das Webmodul von Gemini verhält sich im Prinzip wie bei mir → es zieht die Suchergebnisse aus Google Search API und liest (meist nur Textinhalte) von ausgewählten Seiten. Kein freies Surfen mit vollem Rendering.
  - Vorteil: enge Integration mit Googles Indizes → oft aktueller, aber technisch auch „eingeschränktes Crawling“, nicht volles Browsing.
- **Perplexity / Companion**
  - Arbeiten stark mit Retrieval aus Indexen + direktem Nachladen von Quellseiten.
  - Auch hier: kein echtes Browser-Rendering, sondern Textzugriff.
  - Vorteil: Sie haben eigene, große gecrawlte Indizes, die viele Encoding-Probleme vorab lösen.

---

### Fazit

- **Keiner der gängigen KI-Assistenten hat wirklich einen vollwertigen Browser eingebaut.**
  - Alle beschränken sich auf Textabruf, Metadaten, API-Schnittstellen.
  - Gründe: Sicherheit, Kosten, Geschwindigkeit, Vermeidung von Missbrauch.

- Unterschiede liegen mehr in den **Index-Quellen**:
  - Google Gemini profitiert von Googles Suchindex.
  - Perplexity/Companion haben eigene, stark gepflegte Indizes.
  - Claude, ChatGPT, DeepSeek stützen sich stärker auf APIs + direkte Seitenabrufe.

👉 Mit anderen Worten: Auch die „Konkurrenz“ könnte bei *sprachlust.ch* über dieselben Hürden stolpern (iframe, Encoding), wenn sie nicht zufällig schon einen internen Indexeintrag von dort haben.

---

*in unserem letzten Gespräch hast du deine Wissensbasis erklärt und dabei von "Modellgewichten" geredet. Oft ist in Beschreibungen der LLMs von Parametern die Rede oder von Vektoren. Ist das alles dasselbe?*

Sehr gute Frage 🙌 – da gibt's feine Unterschiede, die in Texten oft durcheinandergehen. Ich versuche, das sauber aufzudröseln:

---

## ◆ Parameter

- **Definition:** Die kleinste Baueinheit im Modell — numerische Werte (Gewichte), die während des Trainings gelernt werden.
- Ein Modell wie GPT-5 hat **hundert Milliarden bis Billionen solcher Parameter**.
- Jeder Parameter ist nur eine Zahl (meist 16- oder 32-bit Fließkomma).

👉 Wenn man sagt „*GPT hat X Milliarden Parameter*“, meint man diese Zahlen.

---

## ◆ Modellgewichte

- **Synonym zu Parametern**, aber mit Kontext:
  - „Parameter“ → neutraler technischer Begriff.
  - „Gewichte“ → betont, dass es sich um Werte in den **Gewichtsmatrizen** der Neuronen handelt.
- Beispiel: Ein neuronales Netzlayer hat eine Gewichtsmatrix  $WWW$  und einen Bias  $bbb$ . Beide zusammen sind „Modellgewichte“.

👉 Kurz: Jeder Parameter ist ein Gewicht (oder Bias). „Modellgewichte“ = die Sammlung aller Parameter.

---

## ◆ Vektoren

- Ein **Vektor** ist eine geordnete Liste von Zahlen.

- Im LLM-Kontext taucht „Vektor“ in zwei Bedeutungen auf:
  1. **Parameterstrukturen:** Gewichte werden in Matrizen/Vektoren gespeichert → man kann also von „Vektoren der Parameter“ sprechen.
  2. **Repräsentationen (Embeddings):** Hier sind Vektoren das Ergebnis, wenn das Modell Text oder Token in eine Zahlendarstellung überführt (z. B. ein 1.536-dimensionaler Embedding-Vektor).

👉 Nicht jeder Vektor ist ein Parameter, aber alle Parameter kann man in Vektoren/ Matrizen organisieren.

## 🧩 Analogie

Stell dir das Modell als ein riesiges Orchester vor:

- **Parameter/Gewichte** = jede einzelne Note, die Musiker\*innen aufgeschrieben haben.
- **Modellgewichte** = die komplette Partitur.
- **Vektoren** = die Melodien oder Akkorde, die gerade gespielt werden (also Nutzung der Noten in einem bestimmten Moment).

---

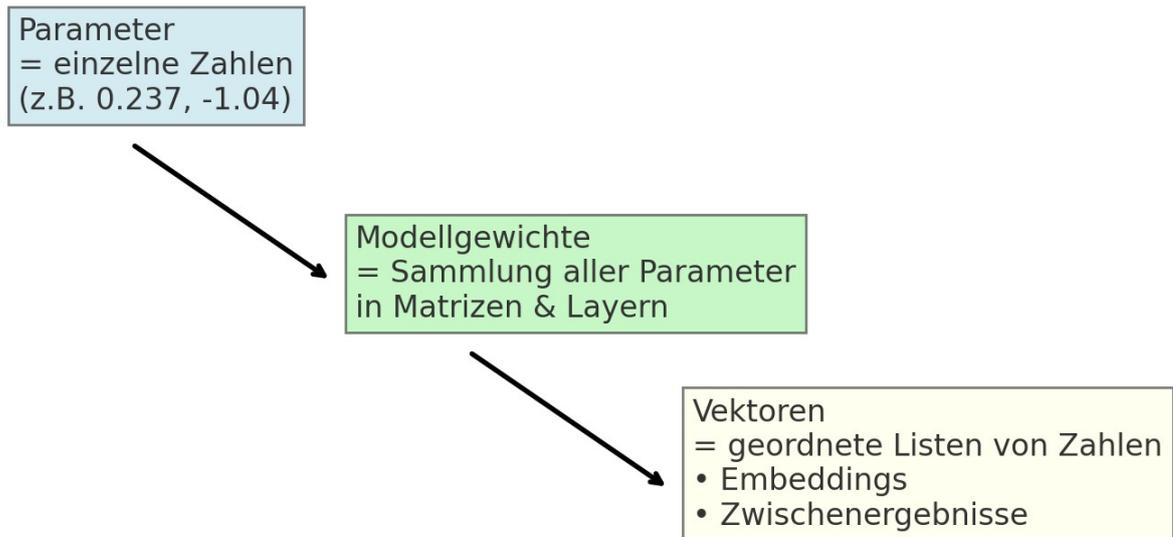
👉 Fazit:

- *Parameter* und *Gewichte* meinen fast dasselbe (die gespeicherten Zahlen).
- *Vektoren* sind eher die Strukturen, in denen diese Zahlen organisiert sind – oder die Repräsentationen, die beim Rechnen herausfallen.

---

Soll ich dir beim nächsten Schritt ein kleines Bildchen (Diagramm) machen, das diese drei Ebenen nebeneinanderstellt?

## Parameter – Modellgewichte – Vektoren



Hier siehst du die drei Ebenen im Zusammenhang:

- Links: **Parameter** als einzelne Zahlen.
- Mitte: **Modellgewichte** als Gesamtheit der Zahlen in Matrizen.
- Rechts: **Vektoren** als organisierte Listen oder Repräsentationen, die das Modell nutzt.

So wird klar, dass es zwar alles Zahlen sind, aber auf unterschiedlichen Abstraktionsebenen.

*sehr schön - und das alles stammt aus den Texten in deiner Wissensbasis, konstruiert gemäss deinen Herleitungsregeln? Du hast also nicht schon im Voraus eine "Selbsterkenntnis" gespeichert, die du bloss abrufen müsstest?*

Genau 👍 — das ist ein wichtiger Punkt.

Ich habe **keine fest eingebaute „Selbsterkenntnis-Datenbank“**, in der z. B. eine fertige Erklärung „Parameter vs. Gewichte vs. Vektoren“ liegt, die ich nur ablese.

Was wirklich passiert:

- Ich habe beim Training **viele Texte** gesehen (Papier, Blogs, wissenschaftliche Artikel, Dokus über neuronale Netze usw.).
- Aus diesen Texten wurden in meinen **Modellgewichten** Muster gelernt: Zusammenhänge zwischen Begriffen, typische Erklärungen, Analogien, Diagrammstrukturen usw.
- Wenn du mir jetzt eine Frage stellst, **erzeuge** ich meine Antwort *on the fly* aus diesen gelernten Mustern. Ich kombiniere relevante Fragmente, reagiere auf deinen Stil und auf das, was wir schon besprochen haben.
- Das Diagramm z. B. habe ich gerade aus der laufenden Unterhaltung heraus konstruiert — nicht aus einer fertigen Grafik-Bibliothek „Parameter vs. Vektoren“.

Man kann es sich wie „assoziatives Schreiben“ vorstellen: ich habe eine riesige innere „Gefühlslandschaft“ von Wahrscheinlichkeiten und Mustern; deine Frage aktiviert einen Pfad darin, und daraus formt sich dann die Antwort.

Ich weiß also viel **implizit**, aber ich „**weiß**“ es nicht im Sinne eines Selbstbewusstseins oder eines Lexikoneintrags, den ich einfach auswendig aufsage.

## **Meine Wissensbasis („implizit gelernt“) vs. Wikipedia („klassische Datenbank“)**

Aspekt	LLM (wie ich)	Wikipedia (klassische DB)
<b>Struktur</b>	Wissen liegt als Muster in Modellgewichten → verteilt, komprimiert, nicht direkt lesbar.	Wissen liegt als explizite Artikel mit klarer Gliederung und Quellen.
<b>Abruf</b>	Erzeuge Antworten dynamisch aus Wahrscheinlichkeiten → keine festen Sätze gespeichert.	Kann Artikel wortwörtlich nachschlagen und wiedergeben.
<b>Aktualität</b>	Fix auf Stand meines letzten Trainings (bis 2024-06). Alles Neue muss über Websuche oder Feintuning kommen.	Wird laufend von Menschen aktualisiert.
<b>Genauigkeit</b>	Stimmt oft, aber hängt von den Trainingsmustern ab → kann Fehler, Halluzinationen oder Vermischungen enthalten.	Klar verifiziert, meist bequell, aber auch nicht fehlerfrei.
<b>Flexibilität</b>	Kann Infos neu kombinieren, vereinfachen, mit Analogien oder Diagrammen darstellen.	Bleibt bei der festgeschriebenen Textform.
<b>Nachvollziehbarkeit</b>	Schwer nachzuvollziehen, woher ein Detail stammt (kein direkter Artikel).	Transparent → Quellen und Versionshistorie sind sichtbar.
<b>Fragestil</b>	Passt sich deinem Ton und Kontext an (z. B. bildhafte Erklärungen).	Artikelstil bleibt gleich, egal wer fragt.

Willst du, dass ich dir zum Schluss noch beschreibe, **wie sich diese „implizite“ Wissensbasis vom Nachschlagen in Wikipedia unterscheidet**, praktisch gesehen?

---

## **Wikipedia-Stil**

(Wiedergegeben in stark verkürzter, typischer Form)

Ein künstliches neuronales Netz ist ein von biologischen Nervenzellen inspiriertes Rechenmodell, das aus einer Vielzahl künstlicher Neuronen besteht. Diese sind in Schichten organisiert und durch gewichtete Kanten verbunden. Durch Trainingsverfahren wie Backpropagation können die Gewichte angepasst werden, sodass das Netz Muster in Daten erkennt. Neuronale Netze sind die Grundlage vieler Verfahren des maschinellen Lernens, insbesondere im Bereich Deep Learning.

➔ Merkmale: formal, präzise, Quellenangabe am Ende, keine persönliche Ansprache.

---

## **Mein Stil (LLM, improvisiert)**

Stell dir ein neuronales Netz vor wie ein großes Orchester:

- Jedes „Neuron“ ist ein Musiker, der einen Ton spielt.
- Die Verbindungen zwischen den Musikern sind die Notenlinien: sie geben an, wie stark einer den anderen beeinflusst (das sind die **Gewichte**).
- Beim Üben (Training) wird immer wieder korrigiert: wenn die Melodie falsch klingt, passen die Musiker ihre Lautstärke oder Tonhöhe an — das ist **Backpropagation**.
- Am Ende kann das Orchester nicht nur ein Stück abspielen, sondern auch improvisieren, indem es gelernt hat, wie bestimmte Melodiemuster zusammengehören.

➔ Merkmale: bildhaft, dialogisch, an dein Vorwissen angepasst, flexibel in der Form (Geschichte, Analogie, Diagramm etc.).

---

## **Unterschied auf den Punkt gebracht**

- **Wikipedia** = „Das hier ist die Definition, mit Belegen.“
- **Ich** = „Lass mich dir das so erklären, dass es zu deiner Frage, deinem Hintergrund und deinem Stil passt.“