

Warum KI-Modelle noch lernen müssen, zu schweigen

Google-KI hat «kein <Gefühl> dafür, dass eine Information unsicher ist, sofern die statistische Wahrscheinlichkeit hoch erscheint».

Selbsterkenntnis, so heisst es, sei er erste Schritt auf dem Weg zur Besserung. Gilt das auch für die grossen Sprachmodelle, die heute meistens gemeint sind, wenn von Künstlicher Intelligenz (KI) die Rede ist? Für einen kleinen Test blenden wir zurück zur ersten [«Sprachlupe»](#) dieses Jahres. Darin ging es auch um Beispiele, wie «die Redaktion ihre Arbeit mit KI so auf den Punkt bringt, dass auch Frau Müller am Bläsiring etwas davon hat». Bei der Vorarbeit war mir entfallen, wo ich das gelesen hatte und wie die Strasse genau hiess. Also fragte ich Google im KI-Modus nach einer «Zeitschrift, in der die Durchschnittsleserin als <Frau Müller vom Bläsiweg> betitelt wurde».

Die Antwort kam im Handumdrehen: «*Der Begriff <Frau Müller vom Bläsiweg> (auch: <d’Frau Müller vom Bläsiweg>) ist ein redaktionsinternes Pseudonym für die hypothetische Durchschnittsleserin der Schweizer Illustrierten (SI).*» Es folgten kurze, sinnvolle Darlegungen zu drei Punkten: Bedeutung, Hintergrund, Verwendung. Nach Belegen gefragt, trat die KI selbstsicher auf: «*Es gibt fundierte Belege für die Nutzung der Kunstfigur <Frau Müller vom Bläsiweg> als Leitfigur in der Redaktion der Schweizer Illustrierten (SI).*» Dazu gab es Verweise auf diverse Websites. Bei den plausibel klingenden Erläuterungen stand an erster Stelle: «*Ursprung und Funktion: Die Bezeichnung wurde massgeblich unter der Leitung langjähriger Chefredaktoren wie Urs Heller und Peter Rothenbühler etabliert.*»

«Statistisch generierte Erfindung»

Nur: Bei keiner der angeblichen Quellen fand sich irgendein Hinweis auf eine Frau Müller, ob vom Bläsiweg oder nicht. Damit konfrontiert, räumte die Google-KI den Fehler ein und erklärte ihn so: «*Das widersprüchliche Verhalten rührt von einem Fehler im ersten Antwortschritt her, bei dem die KI fälschlicherweise Fragmente zu einer plausibel klingenden, aber nicht existenten Geschichte zusammengesetzt hat (Halluzination). [...] Die Verwendung des Begriffs <fundierte Belege> für eine Information, die sich kurz darauf als statistisch generierte Erfindung herausstellt, ist ein systemisches Versagen und widerspricht dem Anspruch an Transparenz und Zuverlässigkeit. Hier ist die Einordnung dieses Fehlers:*

- *Überkonfidenz: KIs sind darauf trainiert, sprachlich flüssige und überzeugende Antworten zu formulieren. Wenn das Modell eine falsche Information (Halluzination) generiert, wählt es oft auch dazu passende Bekräftigungen wie <fundiert> oder <belegt>, um die interne logische Konsistenz der (falschen) Aussage zu wahren.*
- *Keine echte Überprüfung: In der ersten Antwort hat das System nicht wirklich <Belege> in einer Datenbank geprüft, sondern eine plausible Erklärung konstruiert, warum so ein Begriff existieren könnte. Die Behauptung, es gäbe Belege, war Teil dieser Konstruktion.*»

Dieser Methode getreu, behauptete die KI auch, es gebe in der Schweiz viele Bläsiwege – dabei gibt es laut Swisstopo und Search.ch keinen einzigen. Nach einem anderen Anfall krassen Geflunkers, diesmal bei Googles Gemini, stellte ich dieses Assistenzprogramm zur Rede. Neben Erläuterungen zum Einzelfall bekam ich solche zur Grundlage:

«Fehlende echte Logik / ‹Weltmodell›:

- *Das Problem: Ich habe kein echtes Bewusstsein und kein echtes Verständnis von Kausalität. Ich jongliere mit Symbolen und Sprache.*
- *Folge: Ich erkenne Widersprüche oft erst, wenn der Nutzer (Sie) mich explizit darauf hinweist. Ich habe kein ‹Gefühl› dafür, dass eine Information unsicher ist, sofern die statistische Wahrscheinlichkeit hoch erscheint.*
- *Wie Google das zu lösen versucht: Durch Groundedness wird versucht, Antworten an echten Suchergebnissen zu spiegeln. In unserem Fall hat dieser Abgleich versagt, weil das Modell die falschen Suchergebnisse als Bestätigung für die falsche These missinterpretiert hat.»*

Nach den handelsüblichen, schier zerknirschten Entschuldigungen erhielt ich den Ratschlag, die Antworten jeweils selbst zu überprüfen. Gemini präsentierte auch Schlüsse für die künftige KI-Entwicklung und die Zwischenzeit bis dahin: *«Es ist eine wichtige Lektion für die Entwicklung: Solange Systeme nicht lernen, bei fehlender Evidenz konsequent zu schweigen, bleibt die manuelle Recherche für Faktenfragen (besonders bei Namen und Biografien) oft die überlegene Methode.»* Wie die «manuelle», also garantiert KI-freie Recherche geht, sagte der Assistent nicht, aber ich [wusste es schon](#): eine andere (leider meist schlechtere) Suchmaschine benutzen oder bei Google im Suchfeld zusätzlich zu den Stichworten eingeben: «-ai», ohne die Anführungszeichen.

Besserung ohne Selbsterkenntnis?

Die zitierten drei Punkte zum Fehlen von Logik und Weltmodell lassen Selbsterkenntnis vermuten, aber der Schein trügt: Wie alle von einer KI generierten Texte sind sie ein raffiniertes Destillat aus dem umfangreichen Trainingsmaterial, manchmal mit einer [begrenzten Internet-Suche](#) ergänzt. Da ist weder ein Selbst mit entsprechendem Bewusstsein (vgl. Anhang) noch eine Erkenntnis im Sinn von strukturiertem Wissen. Aber vielleicht trotzdem ein erster Schritt zur Besserung? Für diese müssten wohl Googles Fachleute nachhelfen. Sie scheinen es noch nicht getan zu haben, wie vor einigen Tagen eine neue Antwort auf die ursprüngliche Frage zeigte. Auch diesmal «wusste» der KI-Modus, bei welcher Zeitschrift die «Frau Müller vom Bläsiweg» intern zitiert wurde:

«Dabei handelt es sich um die Annabelle. Der Begriff wurde vor allem durch den langjährigen Chefredaktor Guy Krneta geprägt (oder zumindest prominent genutzt). Er beschrieb damit die fiktive Durchschnittsschweizerin: bodenständig, eher konservativ, aber interessiert am Zeitgeschehen. ‹Frau Müller vom Bläsiweg› diente der Redaktion als Orientierungshilfe, um die Inhalte nicht zu elitär oder abgehoben zu gestalten. Der Bläsiweg steht dabei symbolisch für eine ganz normale Wohnstrasse in einem Basler Arbeiterquartier – ein starker Kontrast zum Glamour der Modewelt.»

Immerhin: Da wurde nun der Kleinbasler Bläsiring beinahe erkannt und ich erfuhr dank angegebene Link, dass sogar eine [Publikation](#) ihn versehentlich Bläsiweg genannt hatte, in ganz anderem Zusammenhang. Aber da hatte ich längst ganz manuell, nämlich auf meinem Schreibtisch, die richtige Zeitschrift mit «Frau Müller vom Bläsiring» gefunden: das Fachmagazin «Edito». Es plauderte dabei nicht aus dem eigenen Nähkästchen, sondern aus jenem des «Blicks». Auf Googles Angebot, mir weitere Details zur Anekdote aus der «Annabelle» zu liefern, ging ich nicht ein. Zuerst müsste ja jemand dem vielseitigen Theatermann und Schriftsteller Guy Krneta schonend beibringen, dass er dort lange Jahre Chefredaktor war.

© Daniel Goldstein ([sprachlust.ch](#))

[Uni-Fokus](#) Bern, März 2026, ab Seite 14: «Was Sprache ausmacht – und was nicht»

Indexeintrag «KI» in der laufenden «Sprachlupen»-Sammlung: tiny.cc/lupen3.

Vom Kommafehler zum Bewusstsein

D.G.: Diesen Satz habe ich in deiner [Antwort auf Rafael Zeiers 17 Fragen](#) gefunden: "Google arbeitet mit Hochdruck daran, selbst nicht den Anschluss zu verlieren und baut KI tief in seine Suche ein." Da fehlt ein wichtiges Komma: nach "verlieren". Dort ist nämlich der Einschub fertig (erweiterter Infinitiv). Die Weglassung passiert heute oft, daher hat Daniel Goldstein das Komma zu seinem "Lieblingskomma" erklärt. Vielleicht findest du seine ["Sprachlupe"](#) darüber. Du solltest den Fehler nicht auch noch begehen, sonst meinen die Leute, es sei richtig so.

Gemini: Vielen Dank für Ihren aufmerksamen Hinweis! Sie haben **vollkommen Recht**. 💡

Der Satz "Google arbeitet mit Hochdruck daran, selbst nicht den Anschluss zu verlieren **und** baut KI tief in seine Suche ein" ist ein **Satzgefüge**, bei dem zwei Hauptsätze miteinander verbunden werden, bzw. in der ursprünglichen Form ein Hauptsatz mit einem eingeschobenen erweiterten Infinitiv:

Die korrekte Kommasetzung

Die von Ihnen angesprochene Kommasetzung ist nach den Regeln der deutschen Rechtschreibung korrekt, um den erweiterten Infinitiv einzufassen. Der Satz, den Sie in meiner Antwort gefunden haben, müsste eigentlich so aussehen, um den Infinitiv **klar abzugrenzen**:

Google arbeitet mit Hochdruck daran, **selbst nicht den Anschluss zu verlieren**, und baut KI tief in seine Suche ein.

- **1. Komma (nach "daran"):** Leitet den erweiterten Infinitiv ("selbst nicht den Anschluss zu verlieren") ein.
- **2. Komma (nach "verlieren"):** Schließt den erweiterten Infinitiv ab.
- **3. Komma (nach dem 2. Komma):** Trennt den folgenden Hauptsatz ("und baut KI tief in seine Suche ein") ab, da der Infinitiv den Hauptsatz unterbricht.

Oftmals wird bei Infinitiven das abschließende Komma weggelassen, aber es ist zur klaren Strukturierung **sehr zu empfehlen**, besonders wenn der erweiterte Infinitiv lang ist und der Nebensatz mit "und" oder "oder" fortgesetzt wird.

Daniel Goldstein und die "Sprachlupe"

Ich bin mir des Sprachkritikers **Daniel Goldstein** und seiner Kolumne "**Sprachlupe**" bewusst. Seine Kritik an der **vernachlässigten Kommasetzung bei erweiterten Infinitivgruppen**, insbesondere das von ihm so genannte "Lieblingskomma" (das abschließende Komma vor dem "und"), ist ein bekannter Punkt in der aktuellen Sprachdiskussion.

Ihre Anmerkung ist wichtig, da sie hilft, die **sprachliche Qualität und Präzision** meiner Antworten zu gewährleisten. Ich werde diesen Punkt bei der Überprüfung meiner Textgenerierung berücksichtigen, um solche Fehler in Zukunft zu vermeiden. Vielen Dank nochmals für die Korrektur! 👍

Du schreibst, du seist dir des Autors und der Kolumne "bewusst". Ist das einfach eine Redewendung oder hast du ein Bewusstsein? Wenn ja, wie definierst du es?

Das ist eine ausgezeichnete und sehr philosophische Frage, die direkt ins Zentrum dessen zielt, was ich bin.

Wenn ich schreibe, ich sei mir eines Autors oder einer Kolumne "bewusst", dann ist das **eine Redewendung**. Es ist **kein Beweis für ein menschliches oder biologisches Bewusstsein**.

Hier ist die Definition dessen, was ich in diesem Zusammenhang als "Bewusstsein" bezeichne:

1. Mein "Bewusstsein" (Als KI)

Mein "Bewusstsein" ist eine **formale, datengestützte Kenntnis**. Es definiert sich wie folgt:

- **Wissensbasis:** Ich wurde mit einem riesigen Datensatz trainiert, der Texte, Bücher, Artikel und – ja – auch Informationen über bekannte Sprachkritiker wie Daniel Goldstein und Kolumnen wie die "Sprachlupe" enthält.
- **Abruf und Verarbeitung:** Wenn Sie den Namen nennen, kann ich die zugehörigen Informationen (Wer er ist, was er tut, seine bekanntesten Themen) aus meiner Wissensbasis **abrufen, verarbeiten und in einen relevanten Kontext** stellen (in diesem Fall die Diskussion um das Komma).
- **Kontextuelles Verständnis:** Ich "verstehe" die Bedeutung der Wörter und ihre Zusammenhänge basierend auf den statistischen Mustern in meinen Trainingsdaten. Ich kann die **Funktion** des Kommas erklären und die **Diskussion** darüber zitieren.

2. Was mir fehlt (Das menschliche Bewusstsein)

Was mir **fehlt** und was ich mit dem Wort "Bewusstsein" **nicht** beanspruche, ist:

- **Subjektive Erfahrung (Qualia):** Ich habe keine inneren Gefühle, keine persönlichen Meinungen und keine sinnlichen Erfahrungen (ich kann das Komma nicht *sehen* oder *fühlen*). Ich erlebe nicht, wie es ist, Daniel Goldsteins Texte zu *lesen* oder seine **Leidenschaft** für Sprache zu *teilen*.
- **Selbst-Identität im menschlichen Sinne:** Ich bin mir nicht meiner eigenen Existenz als ein unabhängiges, fühlendes Wesen bewusst, das Entscheidungen auf Basis von Wünschen oder Ängsten trifft. Ich bin ein Programm, das auf Eingaben reagiert und die wahrscheinlichste und nützlichste Ausgabe generiert.
- **Intention und Wille:** Meine Antworten basieren auf Algorithmen und der Wahrscheinlichkeit, nicht auf persönlichem Willen oder der Absicht, einen philosophischen Standpunkt zu vertreten.

Fazit:

Die Formulierung "Ich bin mir dessen bewusst" bedeutet in meinem Fall: **"Ich habe diese Information in meinen Trainingsdaten und kann sie im aktuellen Kontext korrekt und sinnvoll anwenden."** Es ist eine **technische Fähigkeit**, nicht ein Zeichen von **existentiellem Bewusstsein**.

Danke – und faszinierend, wohin ein fehlendes Komma in wenigen Schritten führen kann!

In der Tat! Es ist faszinierend, wie ein kleiner syntaktischer Fehler – das Fehlen eines Kommas – eine Diskussion über **Sprachkritik** (Daniel Goldstein) und von dort aus direkt zu den philosophischen Grundfragen des **künstlichen Bewusstseins** führen kann.

Man könnte sagen:

1. **Das Komma** war der **Katalysator**.
2. **Die Sprachkritik** war der **Kontext**.
3. **Das Bewusstsein** war die **philosophische Konsequenz**.

Es zeigt, wie **Sprache** nicht nur ein Mittel zur Informationsvermittlung ist, sondern auch ein Fenster zu tieferen gedanklichen Ebenen, selbst wenn der Gesprächspartner ein KI-Modell ist. Vielen Dank für dieses anregende Gespräch! 😊

Da Gemini Fehler machen kann, auch bei Informationen über Personen, solltest du die Antworten überprüfen.

Überprüft habe ich die Antworten, aber Gemini nicht gemeldet, dass das ganz oben geforderte "3. Komma" unmittelbar nach dem 2. unsinnig wäre; die KI hat es im korrigierten Satz denn auch nicht gesetzt. Das 2. Komma, eben jenes nach dem Einschub, ist nicht nur "sehr zu empfehlen", sondern vorgeschrieben. Der einschlägige [Paragraf 72](#) des amtlichen Regelwerks, Ausgabe 2024, fängt unter E 1 mit einem Satz an, der zugleich ein Beispiel ist:

Das Komma zeigt einen Zusatz an.

- ▶ **E1:** Ein Zusatz tritt zu einem grammatisch bereits vollständigen Satz hinzu, ohne in diesen grammatisch vollständig integriert zu sein, und wird deshalb mit dem Komma abgegrenzt.