

LEARNING NUGGET E-LEARNING & KI 1. OKTOBER 2026

Modell, LLM, Chat, Agent: vier Begriffe, ein Bild

Modell, LLM, Chat und Agent werden oft verwechselt. Ein Schalenbild zeigt, wie die vier Begriffe zusammenhängen und was das für deinen Unterricht heisst.

Von Rico Eberle, E-Learning-Experte für Weiterbildung und Gemeinderat in Dübendorf

RICO EBERLE · LEARNING NUGGET

Modell **LLM**
Chat **Agent**
Vier Begriffe, ein Bild.



Video zum Beitrag

Dieses Learning Nugget gibt es als Video mit Transkript (1:35 Min.). QR-Code scannen oder Link öffnen:

ricoeberle.ch/blog/modell-llm-chat-agent/

Modell, LLM, Chat, Agent: Die vier Begriffe fallen in fast jedem Gespräch über KI, und oft meinen zwei Personen damit nicht dasselbe. Dabei hängen sie einfach zusammen, wie Schalen, eine um die andere. Das Bild ist eine didaktische Vereinfachung, aber es hilft, Werkzeuge einzuordnen und die richtigen Fragen zu stellen.

Im Kern: das Modell

Ganz innen liegt das Modell. Ein Modell ist das Ergebnis des Trainings: eine riesige Datei mit Milliarden von Zahlen, den sogenannten Parametern. Ein LLM, ein Large Language Model, ist eine bestimmte Art von Modell. Es wurde mit sehr viel Text trainiert und sagt Stück für Stück voraus, wie ein Text weitergeht. GPT, Claude, Gemini oder das Schweizer Apertus sind solche Sprachmodelle.



Im Kern liegt das Modell. Ein LLM ist eine Art Modell und bildet keine eigene Schale.

Wie gross solche Modelle sind, zeigt Apertus: EPFL, ETH Zürich und das Schweizerische Supercomputing-Zentrum CSCS haben es 2025 in zwei Grössen veröffentlicht, mit 8 und mit 70 Milliarden Parametern.

Die «Stücke», von denen im Video die Rede ist, heissen in der Fachsprache Tokens: Wörter, Wortteile oder Satzzeichen. Das Video sagt bewusst «Stück» statt «Wort», weil ein Modell nicht Wort für Wort arbeitet. Warum diese Vorhersage manchmal flüssig klingt und trotzdem falsch ist, zeigt der Beitrag [Die KI lügt nicht. Sie rät.](#)

Die erste Hülle: der Chat

Allein tut ein Modell nichts. Es braucht eine Hülle, über die man es nutzen kann. Die erste Hülle ist der Chat: ein Eingabefeld, ein Verlauf, vielleicht ein paar Dateien. Du fragst, das Modell antwortet.

Ein bekannter Name macht den Unterschied greifbar: ChatGPT ist der Chat, das Modell darin heisst GPT. Wenn im Kollegium jemand sagt, «ChatGPT» habe etwas falsch gemacht, lohnt sich deshalb die Rückfrage, ob die Antwort des Modells gemeint ist oder eine Funktion der Oberfläche.

Die zweite Hülle: der Agent

Der Agent gibt dem Modell ein Ziel und Werkzeuge: im Web suchen, Dateien öffnen, Programme ausführen. Und er lässt es in einer Schleife arbeiten: planen, handeln, prüfen, weitermachen, bis die Aufgabe erledigt ist.



Der Agent gibt dem Modell ein Ziel und Werkzeuge und lässt es in einer Schleife arbeiten.

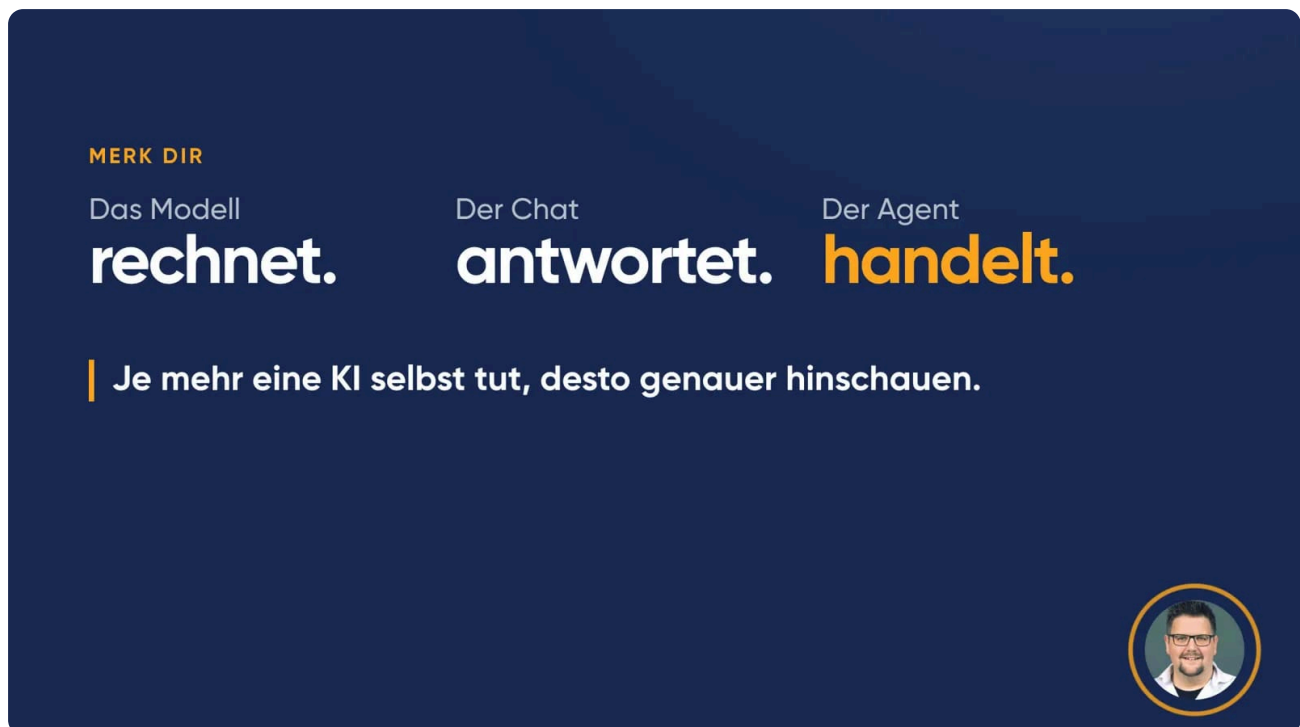
Anthropic beschreibt Agenten in einem Leitfaden von 2024 als Systeme, in denen das Sprachmodell seinen Ablauf und den Einsatz von Werkzeugen selbst steuert. Davon unterscheidet der Leitfaden Workflows, bei denen ein fest programmierter Ablauf vorgibt, wann das Modell welches Werkzeug nutzt. Er nennt auch die Kehrseite: Agenten verursachen höhere Kosten, und Fehler können sich über mehrere Schritte aufschaukeln. Empfohlen werden deshalb ausgiebige Tests in abgeschotteten Umgebungen und passende Leitplanken.

Die Grenzen verschwimmen

Viele Chats können heute auch im Web suchen. Dann arbeiten sie schon ein Stück wie ein Agent: Das Modell entscheidet selbst, ob und wonach es sucht, und baut die Ergebnisse in seine Antwort ein. Das Schalenbild ist deshalb keine scharfe Einteilung, sondern eine Orientierung. Je weiter aussen, desto mehr tut die KI selbst.

Was das für den Unterricht heisst

Der Merksatz aus dem Video fasst das Bild zusammen: Das Modell rechnet, der Chat antwortet, der Agent handelt.



Das Modell rechnet, der Chat antwortet, der Agent handelt.

Für Kursleitende ergeben sich daraus zwei praktische Punkte:

Genau benennen, was gemeint ist. Wer Regeln für KI im Kurs formuliert, sollte sagen, ob ein Chat gemeint ist oder auch Werkzeuge, die selbständig suchen, Dateien bearbeiten oder Programme ausführen.

Genauer hinschauen, je mehr die KI selbst tut. Bei einer Chat-Antwort prüft man einen Text. Bei einem Agenten kommt dazu, welche Schritte er gemacht, welche Quellen er geöffnet und was er verändert hat. Das ist meine Einordnung; sie passt zu Anthropic's Hinweis auf höhere Fehlerfolgen und den Bedarf an Tests und Leitplanken.

Wie ein KI-Assistent mit eigenen Dokumenten arbeitet und was das Sprachmodell dabei wirklich sieht, zeigt der Beitrag [RAG erklärt: Deine KI hat das Handbuch nie gelesen.](#)

Mein Vorschlag

Zeichne in der nächsten Weiterbildung drei Kreise ineinander aufs Flipchart: Modell, Chat, Agent. Lass die Teilnehmenden die KI-Werkzeuge, die sie kennen oder nutzen, auf Haftnotizen schreiben und einem Kreis zuordnen. Spannend sind die Grenzfälle, etwa ein Chat mit Websuche. Frag dort nach: Was tut das Werkzeug selbst, und was muss ich danach prüfen?

Wer die Schalen auseinanderhält, weiss, wo er genauer hinschauen muss.

Häufige Fragen

Was ist der Unterschied zwischen ChatGPT und GPT?

ChatGPT ist der Chat, also die Oberfläche mit Eingabefeld und Verlauf. GPT ist das Sprachmodell darin, das die Antworten berechnet.

Was ist ein LLM?

Ein Large Language Model (LLM) ist ein KI-Modell, das mit sehr viel Text trainiert wurde und Stück für Stück, genau genommen Token für Token, voraussagt, wie ein Text weitergeht. Beispiele sind GPT, Claude, Gemini und das Schweizer Modell Apertus.

Was ist ein KI-Agent?

Ein KI-Agent gibt einem Sprachmodell ein Ziel und Werkzeuge, etwa Websuche, Dateien oder Programme, und lässt es in einer Schleife planen, handeln und prüfen, bis die Aufgabe erledigt ist. Anthropic beschreibt Agenten als Systeme, in denen das Modell seinen Ablauf und den Einsatz von Werkzeugen selbst steuert.

Quellen

1. ETH Zürich: Apertus, ein vollständig offenes, transparentes und mehrsprachiges Sprachmodell (2025)
<https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2025/09/medienmitteilung-apertus-ein-vollstaendig-offenes-transparentes-und-mehrsprachiges-sprachmodell.html>
2. Kalai et al.: Why Language Models Hallucinate (2025)
<https://arxiv.org/abs/2509.04664>
3. Anthropic: Building effective agents (2024)
<https://www.anthropic.com/engineering/building-effective-agents>



ÜBER DEN AUTOR

Rico Eberle

Rico Eberle ist E-Learning-Experte, Betriebsökonom FH und Gemeinderat in Dübendorf. Er präsidiert den Stiftungsrat der Stiftung Weiterbildungskurse Dübendorf (WBK). In den Learning Nuggets erklärt er Forschung zu Lernen, KI und digitaler Souveränität kurz und mit Quellen.

Kontakt: ricoeberle.ch · LinkedIn: [linkedin.com/in/rico-eberle](https://www.linkedin.com/in/rico-eberle)

LIZENZ

Text, Transkript und Video von Rico Eberle stehen unter der Lizenz CC BY 4.0 (ausgenommen Musik und Soundeffekte des Videos). Du darfst die Inhalte teilen und bearbeiten, auch im Unterricht, wenn du die Quelle nennst.
creativecommons.org/licenses/by/4.0/deed.de

ZITIERVORSCHLAG

«Modell, LLM, Chat, Agent: vier Begriffe, ein Bild», Rico Eberle, ricoeberle.ch, 2026,
<https://ricoeberle.ch/blog/modell-llm-chat-agent/>, CC BY 4.0