Rico Eberle · LEARNING NUGGETS

LEARNING NUGGET E-LEARNING & KI 1. OKTOBER 2026

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

Modell, LLM, Chat und Agent liegen wie Schalen ineinander: Das Modell rechnet, der Chat antwortet, der Agent handelt.

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

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/

Ist ChatGPT ein Modell? Nein. ChatGPT ist der Chat, das Modell darin heisst GPT.

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

Im Kern 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.

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» aus dem Video 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 ist der Chat: ein Eingabefeld, ein Verlauf, vielleicht ein paar Dateien. Du fragst, das Modell antwortet.

Die Unterscheidung hat praktische Folgen. Sagt im Kollegium jemand, «ChatGPT» habe etwas falsch gemacht, lohnt sich die Rückfrage: Ist die Antwort des Modells gemeint 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.

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.

Der Leitfaden 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.

Je mehr die KI tut, desto genauer hinschauen

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

Für Kursleitende folgen daraus zwei 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 Anthropics Hinweis auf höhere Fehlerfolgen und den Bedarf an Tests und Leitplanken.

In einer Weiterbildung lässt sich das Bild gut üben. Drei Kreise ineinander aufs Flipchart, beschriftet mit Modell, Chat und Agent. Die Teilnehmenden schreiben die KI-Werkzeuge, die sie kennen, auf Haftnotizen und ordnen sie zu. Interessant sind die Grenzfälle, etwa ein Chat mit Websuche. Dort lohnt die Frage, was das Werkzeug selbst tut und was man danach prüfen muss.

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.

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

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 MICH

Rico Eberle

Ich arbeite im E-Learning, bin Betriebsökonom FH und Gemeinderat in Dübendorf und präsidiere den Stiftungsrat der Stiftung Weiterbildungskurse Dübendorf (WBK). Die Learning Nuggets mache ich allein: Thema, Recherche, Text und Video. Fragen und Widerspruch sind willkommen.

Kontakt: ricoeberle.ch · **LinkedIn:** 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