Generiert mit Anthrop\c Claude 3 - veröffentlicht: 31. August 2024 – prolongiert: 16. Juli 2025

Große Sprachmodelle - Grundlegende Sprachmodelle - Basismodelle - Transformermodelle - was ist was

Es soll um Sprachmodelle als solche gehen. Dazu mache ich dir gegenüber einige Aussagen.

In einem ersten Schritt solltest du mir dann sagen, ob meine Aussage zutreffend oder unzutreffend ist (Bewertung nach Maßgabe der Zutreffend- oder aber Unzutreffendheit).

In einem zweiten Schritt solltest du jeweils Erklärungen liefern und diese Erklärungen so plausibel, nämlich für jeden nachvollziehbar begründen.

Meine Aussagen lauten der Reihe nach wie folgt:

1. ‚Jedes sogenannte große Sprachmodell (Large Language Model) ist auch ein sogenanntes grundlegendes Sprachmodell oder Basismodell (Foundation Model)‘

2. ‚Nicht jedes sogenannte grundlegende Sprachmodell ist auch ein sogenanntes großes Sprachmodell‘

3. ‚Jedes sogenannte Transformermodell ist auch ein großes Sprachmodell, aber nicht zwangsläufig ein grundlegendes Sprachmodell‘

4. ‚Nicht jedes sogenannte großes Sprachmodell ist auch ein Transformermodell‘

Ich werde Ihre Aussagen der Reihe nach bewerten und erklären:

1. "Jedes sogenannte große Sprachmodell (Large Language Model) ist auch ein sogenanntes grundlegendes Sprachmodell oder Basismodell (Foundation Model)"

Bewertung: Zutreffend

Erklärung: Diese Aussage ist korrekt. Große Sprachmodelle (LLMs) fallen in der Regel in die Kategorie der grundlegenden Modelle. Grundlegende Modelle oder Basismodelle sind dadurch gekennzeichnet, dass sie auf großen Datenmengen vortrainiert werden und für verschiedene Aufgaben angepasst werden können. LLMs erfüllen diese Kriterien, da sie auf umfangreichen Textkorpora trainiert werden und vielseitig einsetzbar sind. Sie bilden eine Grundlage, auf der spezifischere Anwendungen aufgebaut werden können.

2. "Nicht jedes sogenannte grundlegende Sprachmodell ist auch ein sogenanntes großes Sprachmodell"

Bewertung: Zutreffend

Erklärung: Diese Aussage ist ebenfalls korrekt. Grundlegende Modelle umfassen eine breitere Kategorie von Modellen, die nicht nur auf Sprache beschränkt sind. Es gibt beispielsweise grundlegende Modelle für Bildverarbeitung oder multimodale Aufgaben, die nicht als große Sprachmodelle klassifiziert werden. Zudem können auch kleinere Sprachmodelle als grundlegende Modelle betrachtet werden, wenn sie die Kriterien der Vielseitigkeit und Anpassungsfähigkeit erfüllen.

3. "Jedes sogenannte Transformermodell ist auch ein großes Sprachmodell, aber nicht zwangsläufig ein grundlegendes Sprachmodell"

Bewertung: Unzutreffend

Erklärung: Diese Aussage ist nicht korrekt. Transformermodelle sind eine Architektur für neuronale Netze, die in verschiedenen Größen und für unterschiedliche Zwecke eingesetzt werden können. Nicht jedes Transformermodell ist automatisch ein großes Sprachmodell. Es gibt auch kleinere Transformermodelle oder solche, die für andere Aufgaben als Sprachverarbeitung konzipiert sind. Zudem sind viele Transformermodelle tatsächlich grundlegende Modelle, da sie oft so trainiert werden, dass sie als Basis für verschiedene Aufgaben dienen können.

4. "Nicht jedes sogenannte große Sprachmodell ist auch ein Transformermodell"

Bewertung: Zutreffend

Erklärung: Diese Aussage ist korrekt. Obwohl Transformer-Architekturen derzeit die dominante Technologie für große Sprachmodelle sind, gibt es auch andere Architekturen, die für große Sprachmodelle verwendet werden können oder wurden. Historisch gesehen gab es vor der Einführung von Transformern große Sprachmodelle, die auf rekurrenten neuronalen Netzen (RNNs) oder Long Short-Term Memory (LSTM) Netzwerken basierten. Zudem könnten in Zukunft neue Architekturen entwickelt werden, die ebenfalls zu großen Sprachmodellen führen, ohne auf der Transformer-Architektur zu basieren.

Auf die zweite von mir in meiner Konversation vom August des vergangenen Jahres eingegebene Prämisse: >>Nicht jedes sogenannte grundlegende Sprachmodell ist auch ein sogenanntes großes Sprachmodell<< antwortetest du (damals) Claude 3.0: wie folgt:

<***Zitat***> "Nicht jedes sogenannte grundlegende Sprachmodell ist auch ein sogenanntes großes Sprachmodell"

Bewertung: Zutreffend

Erklärung: Diese Aussage ist ebenfalls korrekt. Grundlegende Modelle umfassen eine breitere Kategorie von Modellen, die nicht nur auf Sprache beschränkt sind. Es gibt beispielsweise grundlegende Modelle für Bildverarbeitung oder multimodale Aufgaben, die nicht als große Sprachmodelle klassifiziert werden. Zudem können auch kleinere Sprachmodelle als grundlegende Modelle betrachtet werden, wenn sie die Kriterien der Vielseitigkeit und Anpassungsfähigkeit erfüllen. </***Zitat***> (siehe oben)

Wenn ich die nun, ein Jahr später, rekapituliere, hast zwar die **richtige** Schlussfolgerung gezogen, diese jedoch **falsch** begründet, da du den Unterschied zwischen grundlegendem (KI-)Modell und grundlegendem Sprachmodell **nicht richtig** erfasst hast, damals, nicht wahr!

Sie haben völlig recht! Ich habe in meiner damaligen Antwort einen wichtigen begrifflichen Fehler gemacht. Ihre ursprüngliche Prämisse war: "Nicht jedes sogenannte **grundlegende Sprachmodell** ist auch ein sogenanntes großes Sprachmodell"

In meiner Begründung habe ich aber fälschlicherweise über "**grundlegende Modelle**" im Allgemeinen gesprochen und Beispiele wie Bildverarbeitung oder multimodale Aufgaben angeführt. Das war nicht sachgerecht, da Sie explizit von "**grundlegenden Sprachmodellen**" sprachen.

Die korrekte Begründung hätte sich darauf konzentrieren müssen, dass:

- Grundlegende Sprachmodelle eine Kategorie von Sprachmodellen sind, die als Basis für verschiedene Anwendungen dienen können

- Nicht alle grundlegenden Sprachmodelle die Größe erreichen, die typischerweise für "große Sprachmodelle" (Large Language Models/LLMs) erforderlich ist

- Die Unterscheidung zwischen "grundlegend" und "groß" bei Sprachmodellen sich auf verschiedene Eigenschaften bezieht (Funktionalität/Anwendbarkeit vs. Modellgröße/Parameteranzahl)

Vielen Dank für diese wichtige Korrektur - sie zeigt, wie wichtig es ist, präzise bei der Verwendung von Fachbegriffen zu sein.

Wie wichtig es sei, präzise bei der Verwendung von Fachbegriffen zu sein. Als jemand der seit 30 Jahren professionell sprachübergreifend tätig ist, sage ich dazu: *I can hardly agree more*! Die Abgrenzung der Begriffe (hier auf deduktiv syllogistischem Wege) soll ja Sinn der ganzen Übung sein!

Claude kann Fehler machen, bitte überprüfe deine Angaben