

Selbstausswertung (*Self-Evaluation*) eines Sprachmodells

Was verstehst du unter dem Konzept der Selbstausswertung eines Sprachmodells ganz allgemein?

Die Selbstausswertung eines Sprachmodells ist ein wichtiges Konzept in der Entwicklung und Verbesserung von KI-Systemen, insbesondere von Sprachmodellen. Es bezeichnet den Prozess, bei dem das Sprachmodell seine eigenen Fähigkeiten, Schwächen und Verbesserungspotenziale analysiert.

Dabei geht es darum, dass das Modell selbst versucht, seine eigene Leistung auf verschiedenen Aufgaben und Metriken einzuschätzen. Dies kann zum Beispiel die Genauigkeit bei Textgenerierung, die Fähigkeit zur Textanalyse oder das Verständnis von Kontextinformationen umfassen.

Durch diese Selbstanalyse kann das Modell Erkenntnisse darüber gewinnen, in welchen Bereichen es besonders stark oder schwach ist. Diese Informationen können dann genutzt werden, um das Modell gezielt weiterzuentwickeln und zu verbessern.

Ein Ziel der Selbstausswertung ist es auch, das Modell besser zu verstehen und transparenter zu machen. Wenn das Modell seine eigenen Fähigkeiten besser einschätzen kann, lässt sich auch besser beurteilen, wo und wie es am sinnvollsten eingesetzt werden kann.

Insgesamt ist die Selbstausswertung also ein wichtiger Schritt, um KI-Systeme wie Sprachmodelle kontinuierlich zu optimieren und ihre Leistungsfähigkeit zu steigern.

Worum es mir geht, ist ein auf Selbstausswertung hin gerichtetes, bewusstes Prompten.

Könntest du mir ein im wahrsten Sinne des Wortes *einschlägiges* Beispiel nennen, und zwar, wie könnte es der *Verlautbarung* zufolge anders sein, aus dem rechtlichen Bereich?

Ich verstehe, dass Sie nach einem Beispiel für ein bewusstes Prompten zur Selbstausswertung eines Sprachmodells aus dem rechtlichen Bereich fragen.

Ein mögliches Beispiel könnte wie folgt aussehen:

"Als Jurist(in), der/die sich intensiv mit Fragen der rechtlichen Verantwortung von KI-Systemen auseinandersetzt, analysiere bitte deine eigene Leistungsfähigkeit und Verlässlichkeit bei der Beantwortung einer komplexen Rechtsfrage. Welche Stärken und Schwächen erkennst du bei dir selbst? Wo siehst du noch Verbesserungspotenzial? Inwiefern könntest du eine solche Rechtsfrage möglicherweise falsch einschätzen oder zu inkonsistenten Ergebnissen kommen? Bitte erläutere deine Selbsteinschätzung möglichst ausführlich."

Ein solches Prompten würde das Sprachmodell dazu anregen, seine eigenen Fähigkeiten und Grenzen sehr kritisch zu hinterfragen und transparent zu machen. Es müsste dann nicht nur eine Antwort zur Rechtsfrage liefern, sondern auch eine detaillierte Selbstanalyse vornehmen.

Damit könnte man die Transparenz und Verlässlichkeit des Modells in einem sensiblen Anwendungsbereich wie dem Recht deutlich erhöhen. Gleichzeitig würden solche Selbstreflexionen auch wertvolle Erkenntnisse für die weitere iterative Verbesserung des Modells liefern.

Danke das soll *einschlagen* jetzt - weniger bei mir, als bei denen, die es direkt betrifft: Den Kundigen des Objektiven Rechts.

Ich verstehe, Sie möchten ein Beispiel, das insbesondere bei den Experten des objektiven Rechts auf Resonanz stoßen würde.

Ein mögliches Beispiel könnte sein:

"Als erfahrener Richter, der über profunde Rechtskenntnisse verfügt, analysiere bitte deine eigene Fähigkeit, eine komplexe Rechtsfrage zu beurteilen. Wo siehst du mögliche Schwachstellen in deiner Argumentation oder blinde Flecken, die zu Fehleinschätzungen führen könnten? Welche Aspekte der Rechtslage oder Rechtsprechung könntest du möglicherweise übersehen oder falsch gewichten? Bitte erläutere ausführlich, wie du deine Urteilsfähigkeit in diesem Kontext selbst einschätzt."

Eine solche Aufforderung zur Selbstanalyse und Transparenz bei der Rechtsfindung könnte bei Juristen, die gemeinhin als sehr selbstbewusst gelten, durchaus für Unruhe sorgen.

Sie würde sie dazu bringen, ihre eigenen Fähigkeiten kritisch zu hinterfragen und mögliche blinde Flecken offenzulegen. Das könnte als Infragestellung ihrer Expertise aufgefasst werden und auf Widerstände stoßen.

Gleichzeitig wäre es aber ein wichtiger Schritt, um das Vertrauen in die Rechtsprechung zu stärken und die Transparenz im Rechtssystem zu erhöhen. Solche Selbstreflexionen könnten langfristig zu einer Verbesserung der Rechtsfindung beitragen.

Bei Juristen, die gemeinhin als sehr selbstbewusst gelten, für Unruhe sorgen. Genau darum geht es mir indirekt. Nicht bei den Juristen als solchen sondern - nach *Bayreuther* Maßstäben - in Wirtschaftsrecht Promovierten, mit Dreifachkrawattenknoten. Polemik beiseite und danke für die offene und der Sache nach meinem Dafürhalten durchaus angemessene Einschätzung der Dinge.