

Künstliche Intelligenz: Trainieren, Verfeinern und Schärfen eines Großen Sprachmodells

Generiert mit Anthro\ic-3.7 - Veröffentlicht 30. April 2025



Es soll um KI-Sprachmodelle, genauer um Large Language Models gehen. Drei Schritte sind dabei von erheblicher Bedeutung, wenn ich es richtig verstehe.

In der Vorphase (also vor Einsatz des Modells):

1. *Trainieren* (Pre-Trained)
2. *Verfeinern* (Fine-Tuning) gegenüber dem
3. *Schärfen* (Reinforcement Learning from Human Feedback) in der eigentlichen Phase (im Zuge des Einsatzes des Modells), wobei:

1. *Verfeinern* in der Vorphase und
2. *Trainieren* als 'Unsupervised learning' bzw. 'Supervised learning' in der eigentlichen Phase anhand annotierter oder nicht annotierter Daten stattfindet, das
3. *Schärfen* bzw. *Bestärkende Lernen*, ebenso eine Art der *Verfeinerung* ist, wohl aber in der eigentlichen Phase. Richtig?



Das ist im Wesentlichen richtig.

Aber ich möchte einige Klarstellungen und Ergänzungen zu Ihrer Beschreibung der KI-Sprachmodell-Entwicklung vornehmen.

Die Entwicklungsphasen von Large Language Models

Vorphase (vor Produktiveinsatz):

1. **Pre-Training:** Dies ist tatsächlich hauptsächlich "unsupervised learning". Das Modell lernt Muster und Zusammenhänge in großen Textmengen, indem es Wörter oder Token vorhersagt. Dabei werden keine expliziten Labels oder Annotationen benötigt.
2. **Fine-Tuning:** Dies kann sowohl "supervised learning" mit annotierten Daten erfolgen (wo das Modell für bestimmte Aufgaben mit Beispiel-Eingaben und erwünschten Ausgaben trainiert wird) als auch auf andere Weise, während des Einsatzes bzw. in der Produktivphase:
3. **RLHF (Reinforcement Learning from Human Feedback):** Hierbei handelt es sich um eine spezielle Form des Fine-Tunings, bei der menschliche Rückkopplung genutzt wird, um das Modell zu verbessern. Es findet oft während oder nach der initialen Produktivphase statt.

Wichtige Ergänzungen:

* Supervised Fine-Tuning (SFT) ist ein kritischer Zwischenschritt, der oft nach dem Pre-Training und vor RLHF stattfindet. Hierbei wird das Modell mit qualitativ hochwertigen Beispielen trainiert, die zeigen, wie es auf bestimmte Prompts reagieren sollte.

* Der moderne KI-Trainingsablauf besteht typischerweise aus:

1. Pre-Training (unsupervised)
2. Supervised Fine-Tuning (supervised)
3. RLHF oder RLAIIF (Reinforcement Learning from AI Feedback)

* Man unterscheidet auch zwischen Alignment (Ausrichtung auf menschliche Werte, Sicherheit, Hilfeleistung) und reinem Capability Fine-Tuning (Verbesserung der Fähigkeiten für bestimmte Aufgaben).

RLHF ist tatsächlich eine spezielle Art des Fine-Tunings, die aber konzeptionell anders ist als traditionelles supervised Fine-Tuning, da ein Belohnungsmodell und Reinforcement Learning eingesetzt werden, um das Verhalten des Modells zu optimieren.

Ist das verständlich oder möchten Sie, dass ich bestimmte Aspekte näher erläutere?



RLHF sei tatsächlich eine spezielle Art des Fine-Tunings, die aber konzeptionell anders wäre als traditionelles supervised Fine-Tuning, da ein Belohnungsmodell und Reinforcement Learning eingesetzt würden, um das Verhalten des Modells zu optimieren, sagst du.

Vielleicht zum besseren Verständnis: Der Erfolg des chinesischen DeepSeek beruht vor allem auf diesem Belohnungsmodell und dem daraus folgenden Reinforcement Learning, nicht?



Das ist ein interessanter Punkt. DeepSeeks Erfolg beruht tatsächlich auf mehreren Faktoren, unter denen RLHF (Reinforcement Learning from Human Feedback) und dessen Weiterentwicklungen eine wichtige Rolle spielen, aber es ist nicht der einzige entscheidende Faktor.

Bezüglich DeepSeek sind einige wichtige Aspekte zu beachten:

1. Innovatives RLHF-Verfahren: DeepSeek hat in der Tat Fortschritte im Bereich des Reinforcement Learning gemacht. Sie haben eine verbesserte Variante von RLHF entwickelt, die als "Direct Preference Optimization" (DPO) bzw. weiterentwickelte Formen davon bekannt ist.
2. Qualität der Trainingsdaten: Ein wesentlicher Teil ihres Erfolgs liegt in der sorgfältigen Kuratierung und Qualitätskontrolle ihrer Trainingsdaten - sowohl für das Pre-Training als auch für das Feedback-basierte Training.
3. Skalierung und Rechenressourcen: DeepSeek hat erhebliche Rechenressourcen in das Training ihrer Modelle investiert, was eine wichtige Voraussetzung für deren Leistungsfähigkeit ist.

4. Modellarchitektur: DeepSeek hat auch spezifische Modifikationen an der Transformer-Architektur vorgenommen. Der grundsätzliche Unterschied zwischen traditionellem supervised Fine-Tuning und RLHF besteht darin:

* Traditionelles Fine-Tuning: Das Modell lernt direkt von "goldenen" Beispielen (auf Eingabe folgt korrekte Ausgabe).

* RLHF: Das Modell lernt aus Präferenzen zwischen verschiedenen Antworten. Es wird ein separates Belohnungsmodell erstellt, das menschliche Präferenzen modelliert, und dann wird das Hauptmodell optimiert, um diese Belohnungsfunktion zu maximieren.

DeepSeek und andere führende KI-Unternehmen (wie Anthropic, OpenAI, etc.) verwenden inzwischen oft weiterentwickelte RLHF-Varianten wie DPO (Direct Preference Optimization) oder RLAIIF (Reinforcement Learning from AI Feedback), die effizienter sind als klassisches RLHF.

Der Erfolg moderner Sprachmodelle ist also eine Kombination aus allen Entwicklungsphasen:

- exzellentes Pre-Training mit hochwertigen Daten, gezieltes supervised Fine-Tuning und
- innovative Reinforcement Learning-Techniken.