

Im Fokus: DeepSeek

Ausgabe 1/2025 | Autor: Dr.-Ing. Klaus Meffert

Inhalt

Was ist DeepSeek?

Was macht DeepSeek besonders?

Bedeutung von DeepSeek

Ist DeepSeek sicher?

Kontakt zur Expertengruppe "Künstliche Intelligenz"

Der IT Klub Mainz & Rheinhessen

Was ist DeepSeek?

Mit "DeepSeek" wird vor allem das **außergewöhnliche Sprachmodell** namens R1 bezeichnet, das von einer chinesischen Firma namens "DeepSeek" veröffentlicht wurde.

KI-Sprachmodelle werden auch als *LLMs* bezeichnet. LLM steht für "Large Language Model".

Mit R1 ist es DeepSeek gelungen, ein Sprachmodell zu erzeugen, das ähnlich gut wie "ChatGPT" ist.

DeepSeek hat für Aufsehen gesorgt, weil deren App aufgrund der Qualität von R1 innerhalb kürzester Zeit alle Rekorde brach und so maximale Aufmerksamkeit erhielt.

Was macht DeepSeek so besonders?

Aufgrund zahlreicher **herausragender Eigenschaften** hat DeepSeek den KI-Markt durchgewirbelt. Die DeepSeek-App wurde innerhalb kürzester Zeit zu der am meisten heruntergeladenen App. Einige der Gründe für den Erfolg:



Hoher Intelligenzfaktor

DeepSeek-Modelle können eine sehr große Anzahl an Neuronen haben, während sie nur einen Bruchteil davon pro Aufgabenstellung benötigen. DeepSeek hat zudem raffinierte Trainingsmethoden erfunden.



Effizienteres Training und Antwortverhalten

Durch die selektive Aktivierung von Experten reduziert sich der Rechenaufwand erheblich im Vergleich zu Modellen, deren neuronales Netz immer vollständig aktiviert ist.



Geringere Kosten

Durch die selektive Modellarchitektur sind die Kosten für den Betrieb deutlich geringer als bei Modellen wie ChatGPT von OpenAI.



Quelloffen & sicher

Jeder kann zu "OpenAI" werden: DeepSeek-Modelle sind frei verfügbar und können lokal betrieben werden. Die lokale Version kommt ohne jeglichen Datentransfer aus.



Erfolgsrezept veröffentlicht

DeepSeek hat zu allem Überfluss verraten, wie jeder selbst ein KI-Modell erstellen kann, dass ChatGPT Konkurrenz machen kann.

Was bedeutet "DeepSeek" ?

Der Begriff "DeepSeek" wird oft als Synonym für verschiedene Begriffe verwendet, die unterschiedliche Bedeutungen haben.

Begriff	Bedeutung	Eigenschaften
DeepSeek	Chinesische Firma	"KI-Rezept" verschenkt Startup? Geniestreiche
DeepSeek R1	Leistungsfähiges Sprachmodell	Open-Source Sehr leistungsfähig Beste Datensicherheit
DeepSeek App	Anwendung für Smartphones	Keine Datensicherheit Vaporisierte die Börse Basiert auf R1
Destillatmodelle	Aufgeschlaute LLMs mit kompakter Größe	Open-Source Beste Datensicherheit R1 war der Lehrer

Ist DeepSeek sicher?

App für Smartphones (Cloud-Variante)

Die DeepSeek-App gilt als ausgesprochen unsicher. Sie sollte generell nicht verwendet werden, auch nicht für rein private Zwecke.

Open-Source Varianten

Die Open-Source Modelle von DeepSeek, inklusive der handlichen Destillatmodelle, können heruntergeladen und komplett lokal/autark auf eigener Hardware betrieben werden. Sie bieten aufgrund dessen **völlige Datensicherheit**. Ein Datentransfer hin zu Dritten kann ausgeschlossen werden. Somit sind KI-Modelle von DeepSeek in der Open-Source Variante sicherer als Cloud-Lösungen wie die von OpenAI oder Microsoft.

Gespeichertes Wissen

In jedem KI-Modell ist über die Trainingsdaten eine riesige Menge an Weltwissen gespeichert. DeepSeek als chinesische Firma hat aus politischen Motiven einige Fakten verdreht und im angebotenen Flaggship-Modell R1 in fragwürdiger Weise abgespeichert. Mittlerweile gibt es Modelle, die dies korrigiert haben.

Bei KI-Anwendungen, die Firmenwissen befragen, wird bei weit verbreiteten Verfahren wie *RAG* grundsätzlich das Firmenwissen bevorzugt und das chinesische Imperativ unterdrückt und unschädlich gemacht.



KI-unterstützte Erstellung

Teile dieses Dokuments, einschließlich der Illustrationen und Layouts, wurden mit Unterstützung künstlicher Intelligenz erstellt.

Was bedeutet "Mixture of Experts"?

"Mixture of Experts" (MoE) ist eine Architektur für neuronale Netzwerke, bei der mehrere spezialisierte Teilnetzwerke (die "Experten") parallel existieren, während ein Routing-Mechanismus (der "Gatekeeper") entscheidet, welche Experten für die jeweils aktuelle Aufgabe aktiviert werden sollen.

Der Transformer-Ansatz

Transformer ist ein weit verbreiteter KI-Ansatz für "intelligente" Modelle. Bei herkömmlichen Transformer-Modellen werden für jede Befragung alle Parameter des Modells verwendet. MoE-Modelle hingegen teilen ihre Kapazitäten in **mehrere spezialisierte Subnetze** auf, die jeweils für bestimmte Arten von Eingaben oder Aufgaben optimiert sind.

Optimierte Architektur

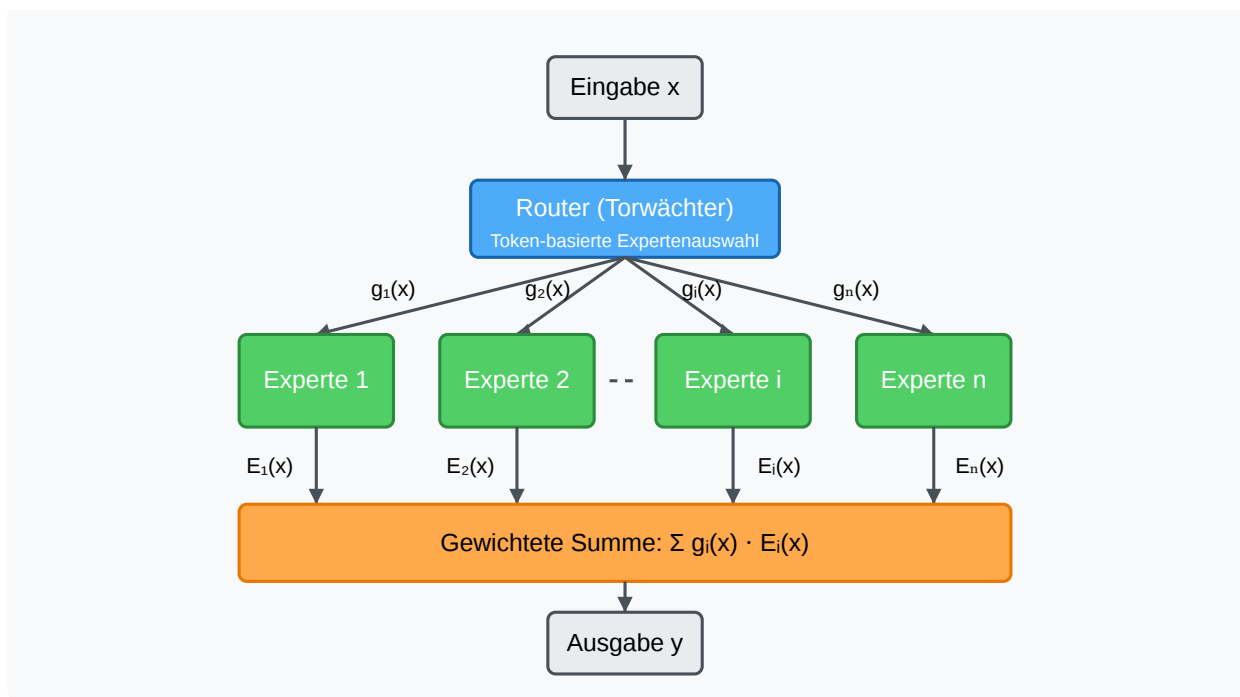
Wenn ein MoE-Modell eine Eingabe erhält, nutzt es zunächst einen **Router** (auch als "Gating Network" bezeichnet). Der Router entscheidet, welche der verfügbaren Experten für diese spezifische Aufgabe am besten geeignet sind. Typischerweise werden dabei nur 1-2 Experten pro Token aktiviert, während die anderen inaktiv bleiben.

Andere MoE-Modelle

DeepSeek hat den MoE-Ansatz nicht etwa erfunden (wohl aber andere Innovationen hervorgebracht). Beispielsweise hat schon das französische Unternehmen Mistral Ende 2023 ein Modell mit dieser Architektur bereitgestellt.

Wie ist DeepSeek R1 aufgebaut?

Die Abbildung illustriert, wie Eingaben von DeepSeek-Modellen verarbeitet werden. Anstatt dass das gesamte elektronische Gehirn eingespannt wird, werden je nach Fragestellung nur wenige Teile davon aktiviert, die sogenannten **Experten**.



Schematische Darstellung einer Mixture-of-Experts Architektur. Der Router weist Eingaben den entsprechenden Experten zu. Die einzelnen Komponenten sind auf der nächsten Seite erklärt.

Diese Architektur bietet mehrere wesentliche Vorteile gegenüber herkömmlichen Modellen mit dichter Struktur und erklärt, warum MoE-Modelle wie DeepSeek-MoE mit vergleichsweise geringem Rechenaufwand sehr leistungsfähig sein können. Das Prinzip ist vergleichbar mit Ihrem Gehirn: Beim Sprechen wird überwiegend Ihr Sprachzentrum beansprucht, und andere Teile werden kaum oder gar nicht aktiv.

Wichtige Komponenten eines MoE-Modells

MoE-Modelle wie DeepSeek R1 nutzen folgenden Innovationen:

1. Router-Netzwerk

Bestimmt, welche KI-Experten für welchen Teil der Eingabe (Token) aktiv werden.

2. Experten-Netzwerke

Spezialisierte neuronale Netze, die jeweils auf bestimmte Aufgaben trainiert sind.

3. Experten-Gewichte

Die vom Router berechneten Gewichte, die bestimmen, wie stark jeder Experte zum Endergebnis beiträgt.

4. Gewichtete Summe

Kombiniert die Ausgaben der verschiedenen Experten basierend auf den vom Router zugewiesenen Gewichten.

Zusammengefasst



Open-Source KI-Modelle werden immer besser und effizienter.



Modelle wie DeepSeek R1 können auf erschwinglicher Hardware selbst betrieben werden.



Lokale Implementierungen verbessern die Ergebnisqualität und sparen bei intensiverer Nutzung Kosten.



Die DeepSeek-App (Cloud) sollte keinesfalls genutzt werden.



Die Anleitung zum Erstellen von "ChatGPT 2.0" ist für jeden öffentlich verfügbar und frei nutzbar.



Open-Source Modelle schaffen Unabhängigkeit von Cloud-Diensten und Apps Dritter.

Die Expertengruppe "Künstliche Intelligenz"

Der IT Klub kann von Hause aus geballte IT-Kompetenz nachweisen. Die KI Expertengruppe unterstützt mit praxisnahen Angeboten und Empfehlungen zu technischen und rechtlichen Aspekten. Maßgeschneiderte Lösungen für intelligente KI-Assistenten und gezielte Weiterbildungen sind im Fokus.

Unser Ziel: Mehrwert für Unternehmen mit Hilfe optimierter und sicherer KI-Lösungen, schnelle Erfolge durch pragmatische Vorgehensweise.

Gerne greifen wir Ihre Themen und Anwendungsfälle auf. Sprechen Sie uns an.



Weiterbildung

Online & vor Ort

Publikationen

Praxisrelevantes Wissen



Dr.-Ing. Klaus Meffert

Ansprechpartner der KI-Expertengruppe. Geschäftsführer der IT Logic GmbH (Idstein). Als Diplom-Informatiker mit KI vertraut, als Herausgeber des Blogs "Dr. DSGVO" und als IT-Sachverständiger auch rechtlichen Fragestellungen zugänglich.

✉ klaus.meffert@it-logic.de ☎ +4.916096897264 × 10¹²



Beratung



Entwicklung



Datensicherheit

Über den IT Klub



Der IT Klub Mainz & Rheinhausen wurde mit dem Ziel gegründet, eine Branchenvertretung für Unternehmen in der Region zu sein. Der IT Klub als Netzwerk bietet seinen Mitgliedern Wissenstransfer, Weiterbildung, Nachwuchsförderung und Standortmarketing.

Unsere regelmäßigen Treffen bieten eine Plattform für fachlichen Austausch, praktische Workshops und spannende Vorträge zu den neuesten Entwicklungen in der IT-Welt. Zu unseren Mitgliedern zählen Unternehmen von klein bis groß. Die Landeshauptstadt Mainz arbeitet sehr eng mit dem IT Klub zusammen.

Wir fördern aktiv den Wissenstransfer zwischen verschiedenen Fachbereichen und Erfahrungsstufen. Durch unsere vielfältigen Aktivitäten schaffen wir Raum für Innovation, Kreativität und berufliche Weiterentwicklung.

90+

Mitglieder

15+

Veranstaltungen
pro Jahr

2

Expertengruppen

Besuchen Sie uns online

 itklub.de