

# 4 Überlegungen für die Wahl des richtigen KI-Modells

Unternehmen stehen unter dem Druck, mittels künstlicher Intelligenz (KI) und generativer künstlicher Intelligenz (gen KI) einen Mehrwert zu erzielen, da sich die Technologie kontinuierlich weiterentwickelt. Ein entscheidender Schritt in diesem Prozess ist die Auswahl des richtigen gen KI-Modells, das bestimmen kann, wie effektiv ein Unternehmen KI zur Erreichung seiner strategischen Ziele einsetzen kann. Diese Checkliste enthält 4 wichtige Punkte, die bei der Auswahl eines KI-Modells zu beachten sind, und beschreibt, wie [Red Hat® Enterprise Linux® AI](#) die Modelle bereitstellen kann, die Sie zum Entwickeln maßgeschneiderter Lösungen für Ihre individuellen Use Cases benötigen.

## 1 Welches Modell in welcher Größe sollte mein Unternehmen verwenden?

Die Auswahl der richtigen Modellgröße für Ihr Unternehmen und dessen spezifischer Use Cases ist ein entscheidender Ausgangspunkt, denn sie kann sich direkt auf die Kosten für den Betrieb des Modells, die erforderlichen Kompetenzen für dessen Management und die Genauigkeit der Ergebnisse auswirken.

Die Verwendung eines größeren Modells mit mehr Parametern kann die Kapazität für das Testen komplexer Szenarien erhöhen. Allerdings benötigen größere Modelle (und die Daten, auf denen sie basieren) mehr Rechenressourcen und eine komplexe Infrastruktur. Dies kann die Betriebskosten erheblich erhöhen und auch mehr Kompetenzen für das Management dieser Modelle erfordern.

Diese Kosten und erforderlichen Kompetenzen können noch mehr zunehmen, wenn Sie ein größeres Modell mit Daten anpassen müssen, die für Ihr Unternehmen, Ihre Use Cases oder Ihre Branche relevant sind.

Ein kostengünstigerer Ansatz ist die Verwendung kleinerer, spezialisierterer Modelle, wie beispielsweise die [Large Language Models \(LLMs\) der Granite-Produktfamilie](#), die in Red Hat Enterprise Linux AI enthalten sind. Sie können anhand Ihrer Unternehmensdaten trainiert werden, um Ihre spezifischen Anforderungen ohne hohen Kosten-, Kompetenz- und Zeitaufwand zu erfüllen.

## 2 Wie transparent ist mein Modell?

Ein weiterer wichtiger Faktor, der die Genauigkeit und Effizienz Ihres gen KI-Modells beeinflussen kann, sind die Daten, die zu dessen Training verwendet werden. Dazu gehören Herkunft, Eigentumsrechte, Qualität, Zugänglichkeit, Zuverlässigkeit, Relevanz, Diversität und Menge der Daten.

Die Verwendung eines Modells, dessen Trainingsdaten nicht transparent sind, setzt Ihr Unternehmen erheblichen Risiken aus. Dazu können ungenaue Ergebnisse aufgrund ungenauer Daten gehören. Noch schwerwiegender sind jedoch die potenziellen rechtlichen Konsequenzen der Verwendung eines Modells, das mit urheberrechtlich geschütztem Material trainiert wurde.

Red Hat empfiehlt die Verwendung von Modellen mit transparenten Datenquellen und Datensätzen, wie beispielsweise die Granite-Modellfamilie, die in Red Hat Enterprise Linux AI enthalten ist. Diese Modelle bieten vollständige Transparenz hinsichtlich der Trainingsdatensätze und Modellinhalte, werden vollständig von Red Hat unterstützt und sind durch Red Hat versichert.

Auf diese Weise kann Ihr Unternehmen seine Initiativen im Bereich generative KI sicher vorantreiben, da Sie die Gewissheit haben, dass das Modell von einer vertrauenswürdigen Quelle stammt, die sicherstellt, dass das Modell unter Berücksichtigung sämtlicher Vorschriften ethisch entwickelt und trainiert wurde und nachweislich sensible Daten schützt.

### 3 Sollte ich Open Source- oder proprietäre Modelle verwenden?

Es gibt verschiedene Arten von Lizenzen für LLMs, darunter Open Source-Lizenzen (wie beispielsweise Apache 2.0, allgemeine öffentliche Lizenzen und andere) sowie kommerzielle Lizenzen, die in der Regel proprietär oder subscriptionsbasiert sind.

Die Entscheidung, ob Sie ein Open Source-Modell oder ein proprietäres Modell wählen, kann einen erheblichen Einfluss auf eine Reihe wichtiger Ergebnisse haben, darunter:

- ▶ Operative Effizienz und Kosten
- ▶ Zulässige Anpassungsgrade
- ▶ Eigentum an Anwendungen und Services, die mit diesem Modell entwickelt wurden, sowie an den Daten, die in dieses Modell eingespeist werden

Proprietäre Modelle können in der Anwendung teurer sein. Sie sind möglicherweise schwieriger für Ihre Use Cases oder Ihre Branche trainier- und anpassbar und können in ihrer Anpassbarkeit eingeschränkt sein. Dazu bieten sie nicht die Effizienz kleinerer Open Source-Modelle, die von der Akzeptanz und den Beiträgen der Community profitieren. Darüber hinaus kann die Nutzung einer kommerziellen Lizenz dazu führen, dass Sie den Eigentümer des Modells entschädigen müssen, wenn Sie eine Dienstleistung oder Anwendung verkaufen möchten, die Sie unter Verwendung des Modells erstellt haben, oder dass Sie ihm das Eigentum an den Daten übertragen müssen, die in das Modell eingespeist wurden.

Die über Red Hat Enterprise Linux AI angebotenen Open Source-lizenzierten LLMs der Granite-Produktfamilie bieten operative und finanzielle Effizienz und ermöglichen Ihnen die freie Nutzung, Änderung und Distribution von Modellen, während Sie gleichzeitig die vollständigen Eigentumsrechte an Ihren Daten behalten.

### 4 Wie sollte ich mein Modell trainieren?

Durch individuelles Anpassen Ihres Modells können Sie die Anforderungen spezieller Use Cases effizient, präzise und kostengünstig erfüllen. Die Art und Weise, wie Sie das Modell an Ihre Anforderungen anpassen, kann sich erheblich auf die Effizienz und Effektivität der Modelloptimierung auswirken.

Red Hat Enterprise Linux AI bietet über eine unterstützte, auf dem Lifecycle basierte Distribution von [InstructLab](#) leicht zugängliche Tools zur Modellanpassung.

Die InstructLab Trainingsmethode für LLMs geht auf die Herausforderungen bei der Entwicklung dieser Modelle und die Skalierungsprobleme ein, die bei der herkömmlichen LLM-Schulung auftreten. Dies wird durch einen taxonomiegesteuerten Prozess zur Generierung synthetischer Daten und ein mehrphasiges Tuning Framework zur Verbesserung der Modell-Performance sowie durch eine Open Source Community erreicht. Dadurch wird es Entwicklerinnen und Entwicklern ermöglicht, gemeinsam neue Kompetenzen und Kenntnisse zu den Modellen beizusteuern.

InstructLab Tools für die Modellanpassung

- ▶ bieten Ihnen eine kosteneffiziente Lösung zur Verbesserung der Ausrichtung von LLMs
- ▶ fördern Verbesserungen an Open Source-Modellen mit einem Community-Ansatz
- ▶ demokratisieren den Prozess der Anpassung von Modellen mit individuellen Datensätzen durch leicht zugängliche Tools, Training und Community-Wissen
- ▶ ermöglicht Ihnen, das vollständige Eigentum an Trainingsdatensätzen und Modell-IP zu behalten

#### Mehr erfahren

[In diesem E-Book](#) erfahren Sie, wie Red Hat Sie beim Einstieg in die KI-Innovation unterstützen kann.

#### Einstieg

[Lassen Sie sich vom fachkundigen Red Hat Team](#) beraten, wie Red Hat Ihr Unternehmen beim Einstieg in die KI unterstützen kann.



#### Über Red Hat

Red Hat unterstützt Kunden dabei, ihre Umgebungen zu standardisieren, cloudnative Anwendungen zu entwickeln und komplexe Umgebungen mit [vielfach ausgezeichnetem](#) Support, Training und Consulting Services zu integrieren, zu automatisieren, zu sichern und zu verwalten.

**f** facebook.com/redhatinc  
**X** @RedHatDACH  
**in** linkedin.com/company/red-hat

**EUROPA, NAHOST,  
UND AFRIKA (EMEA)**  
00800 7334 2835  
de.redhat.com  
europe@redhat.com

**TÜRKEI**  
00800 448820640

**ISRAEL**  
1 809 449548

**VAE**  
8000-4449549