

Basismodellplattform für generative KI

Entscheidende Vorteile

- ▶ Update und Verbesserung von Large Language Models (LLMs) durch Nutzende mit InstructLab
- ▶ Sicherer und geschützter Abgleich von LLMs mit proprietären Daten zum Anpassen der Modelle an Ihre Geschäftsanforderungen
- ▶ Schneller Start mit generativer künstlicher Intelligenz (gen KI) und rasche Ergebnisse mit einer bewährten, sicherheitsorientierten Red Hat® Enterprise Linux® Plattform
- ▶ Verpackt als startfähiges Container Image von Red Hat Enterprise Linux für Installation und Updates

Produktübersicht

Red Hat Enterprise Linux AI ist eine unternehmensgerechte Basismodellplattform für gen KI zum Entwickeln, Testen und Bereitstellen von LLMs für Business Use Cases im Zusammenhang mit generativer KI.

In Red Hat Enterprise Linux AI sind folgende Komponenten vereint:

- ▶ Open Source-LLMs der Granite-Familie
- ▶ InstructLab-Tools für die Modellanpassung, die einen von der Community gesteuerten Ansatz für das Fine Tuning von LLMs bieten
- ▶ Ein startfähiges Image von Red Hat Enterprise Linux, zusammen mit gen KI-Libraries und Abhängigkeiten wie PyTorch und AI Accelerator Driver-Software für NVIDIA, Intel und AMD
- ▶ Unternehmensgerechter technischer Support und eine Regelung zum geistigen Eigentum der Modelle, bereitgestellt von Red Hat
- ▶ Mit Red Hat Enterprise Linux AI erhalten Sie die bewährte Red Hat Enterprise Linux Plattform, ergänzt durch die erforderlichen Komponenten, damit Sie direkt mit gen KI starten und erste Ergebnisse erzielen können.

Die Zukunft von KI ist offen und transparent

Red Hat Enterprise Linux AI enthält einen Teil der Open Source-Sprache Granite und Code-Modelle, die vollständig von Red Hat unterstützt werden. Die quelloffenen Granite-Modelle bieten Organisationen kosten- und performanceoptimierte Modelle, die auf eine Vielzahl von Use Cases für gen KI abgestimmt sind. Die Granite-Modelle wurden unter der Apache 2.0-Lizenz veröffentlicht. Zusätzlich zu den als Open Source verfügbaren Modellen sind auch die für das Training der Modelle verwendeten Datensätze transparent und offen.

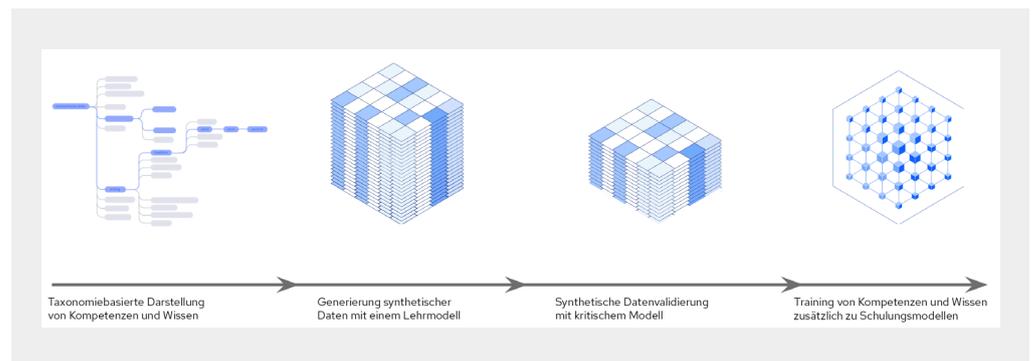
Tabelle 1. Granite-Modelle in Red Hat Enterprise Linux AI

IBM Granite-Sprachmodelle	Granite-7B-Starter Granite-7B-RedHat-Lab
IBM Code-Modelle	Granite-8B-Code-Instruct Granite-8B-Code-Base

Zugängliches Training für gen KI-Modelle für beschleunigte Wertschöpfung

Zusätzlich zu den quelloffenen Granite-Modellen enthält Red Hat Enterprise Linux AI auch InstructLab-Tools zur Modellanpassung, die auf der LAB-Technik (Large Scale Alignment for chatBots) basieren. Mit InstructLab können Teams innerhalb von Organisationen ihre Kompetenzen und ihr Wissen effizient in die LLMs einbringen und diese Modelle an die spezifischen Anforderungen ihres Unternehmens anpassen.

- ▶ Kompetenz: Eine Fähigkeits-Domain, die einem Modell beibringen soll, wie eine bestimmte Aufgabe auszuführen ist. Kompetenzen werden in 2 Kategorien eingeteilt.
 - ▶ Kompositorische Kompetenzen
 - ▶ Lassen KI-Modelle bestimmte Aufgaben oder Funktionen ausführen.
 - ▶ Sind entweder „geerdet“ (beinhalten Kontext) oder „ungeerdet“ (beinhalten keinen Kontext).
 - ▶ Ein Beispiel für „geerdet“ ist das Hinzufügen einer Kompetenz, um einem Modell die Fähigkeit zu verleihen, eine mit Markdown formatierte Tabelle zu lesen.
 - ▶ Ein Beispiel für „ungeerdet“ ist das Hinzufügen einer Kompetenz, um dem Modell das Reimen beizubringen.
 - ▶ Basiskompetenzen
 - ▶ Kompetenzen wie Mathematik, logisches Denken und Programmieren.
- ▶ Wissen: Daten und Fakten, die einem Modell zusätzliche Daten und Informationen zur genaueren Beantwortung von Fragen liefern.



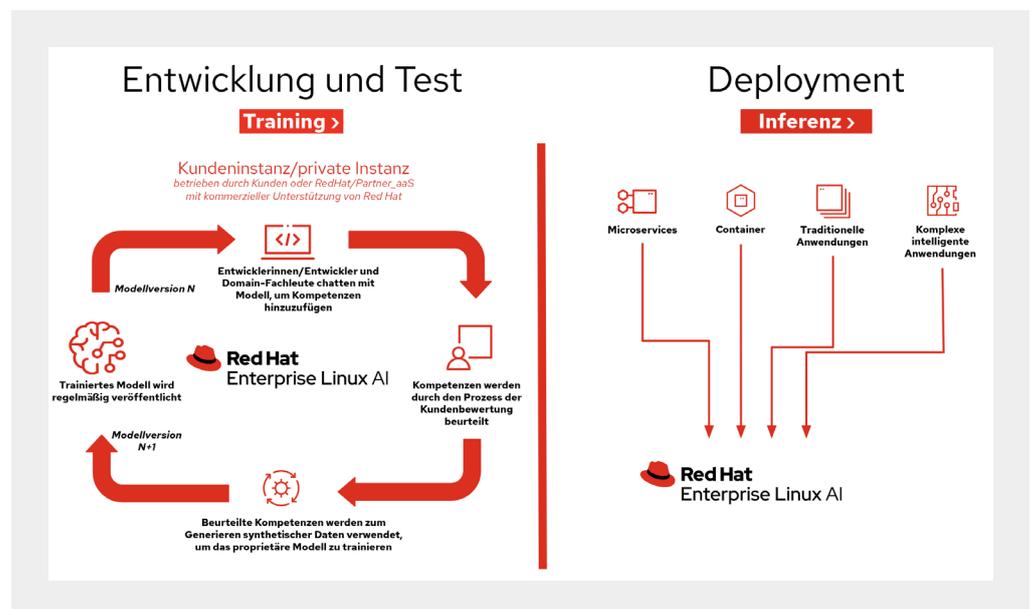
Die obige Abbildung zeigt den Fine Tuning Workflow des InstructLab-Modells.

1. Beiträge zu Kompetenzen und Wissen werden in einem taxonomiebasierten Daten-Repository gespeichert.
2. Unter Verwendung der Taxonomiedaten wird eine erhebliche Menge synthetischer Daten generiert, um einen ausreichend großen Datensatz zu erstellen, mit dem sich ein LLM erfolgreich aktualisieren und ändern lässt.
3. Die synthetische Datenausgabe wird von einem kritischen Modell überprüft, validiert und bereinigt.
4. Das Modell wird mit synthetischen Daten trainiert, die auf manuellen Eingaben von Menschen basieren.

InstructLab ist für Entwicklungsteams und Domain-Fachleute zugänglich, denen möglicherweise das notwendige fachliche Know-how im Bereich Data Science fehlt, das normalerweise für das Fine Tuning von LLMs erforderlich ist. Mit der InstructLab-Methode können Teams ihrem gewählten Modell für das Training Daten oder Kompetenzen hinzufügen, die speziell auf die Anforderungen von Use Cases in Unternehmen zugeschnitten sind. So können sie kollaborativ arbeiten und die Wertschöpfung beschleunigen.

Standortunabhängiges Training und Deployment

Red Hat Enterprise Linux AI unterstützt Unternehmen dabei, den Prozess vom Proof of Concept bis zum produktionsreifen serverbasierten Deployment zu beschleunigen, indem es die erforderlichen Tools bereitstellt und die Möglichkeit bietet, diese Modelle dort zu trainieren, abzustimmen und zu implementieren, wo die Daten gespeichert sind, und zwar in der gesamten Hybrid Cloud. Die so bereitgestellten Modelle können dann von verschiedenen Services und Anwendungen in Ihrem Unternehmen genutzt werden.



Für Organisationen, die dafür bereit sind, bietet Red Hat Enterprise Linux AI auch eine Schnittstelle zu [Red Hat OpenShift® AI](#) für das Training, Tuning und die Bereitstellung dieser Modelle in großem Umfang in einer verteilten Cluster-Umgebung unter Verwendung derselben Granite-Modelle und desselben InstructLab-Ansatzes, die auch beim Deployment von Red Hat Enterprise Linux AI verwendet werden.

Features und Vorteile

Features	Vorteile
Vollständig unterstützte Sprach- und Code-Modelle von Granite, als Open Source-Version unter der Apache 2.0-Lizenz bereitgestellt	Quelloffene und transparente LLMs sowie offen zugängliche Trainingsdaten verbessern die Datentransparenz und beseitigen ethische Bedenken hinsichtlich Dateninhalten und -quellen, wodurch letztlich das allgemeine Geschäftsrisiko gesenkt wird.

Features	Vorteile
Modell-IP-Freistellung für Granite-Modelle	Die Freistellung für die Granite-Modelle in Red Hat Enterprise Linux AI spiegelt das starke Vertrauen von Red Hat und IBM in die gründliche Entwicklung und Prüfung dieser Modelle wider. Diese Freistellung bietet Kunden mehr Sicherheit und ermöglicht es ihnen, die Granite-Modelle mit größerem Vertrauen in das Engagement von Red Hat für ihren Erfolg zu nutzen.
LLM-Anpassungstools von InstructLab für skalierbares und zugängliches Fine Tuning von Modellen	InstructLab bietet eine zugängliche Methode zum Fine Tuning von LLMs, wodurch der Bedarf an tiefgreifendem fachlichem Know-how im Bereich Data Science verringert wird und unterschiedliche Rollen innerhalb Ihrer Organisation einen Beitrag leisten können. So kann Ihr Unternehmen gen KI einführen und damit die Amortisationszeit verkürzen sowie die Kapitalrendite maximieren.
Optimierte, startfähige Modell-Runtime-Instanzen	Red Hat Enterprise Linux AI wird als startfähiges Container Image bereitgestellt, eine Deployment-Methode, die als Image-Modus für Red Hat Enterprise Linux bezeichnet wird. Diese Technologie reduziert die Komplexität von Installation, Konfiguration und Aktualisierung und ermöglicht so einen einfachen Einrichtungs- und Änderungsmanagementprozess.
Abhängigkeiten des gen KI-Pakets und der Softwaretreiber für die KI-Hardware	Starten Sie gen KI sofort mit umfassenden Tools, einschließlich wichtiger Pakete und Treiber wie PyTorch, vLLM und NVIDIA-Treiber, damit Sie vom ersten Tag an für Ihre geschäftlichen Use Cases mit gen KI gerüstet sind.



Über Red Hat

Red Hat, weltweit führender Anbieter von Open Source-Softwarelösungen für Unternehmen, folgt einem communitybasierten Ansatz, um zuverlässige und leistungsstarke Linux-, Hybrid Cloud-, Container- und Kubernetes-Technologien bereitzustellen. Red Hat unterstützt Kunden bei der Entwicklung cloudnativer Applikationen, der Integration neuer und bestehender IT-Anwendungen sowie der Automatisierung, Sicherung und Verwaltung komplexer Umgebungen. [Als bewährter Partner der Fortune 500](#)-Unternehmen stellt Red Hat [vielfach ausgezeichnete](#) Support-, Trainings- und Consulting-Services bereit, die unterschiedlichen Branchen die Vorteile der Innovation mit Open Source erschließen können. Als Mittelpunkt eines globalen Netzwerks aus Unternehmen, Partnern und Communities unterstützt Red Hat Unternehmen bei der Steigerung ihres Wachstums und auf ihrem Weg in die digitale Zukunft.

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