

Erste Schritte mit KI für Unternehmen: Ein Guide für den Einstieg



Red Hat

Inhaltsverzeichnis



Einleitung

Seite 3

Kapitel 1

Der Aufstieg der KI

Seite 6

Kapitel 2

Auswählen des richtigen KI-Modells

Seite 9

Kapitel 3

Was Sie für den Einstieg benötigen

Seite 12

Kapitel 4

Einführen und Skalieren mit Red Hat

Seite 16

Mehr erfahren

Sind Sie bereit für den nächsten Schritt bei der Einführung von KI?

Seite 20

Einleitung

Unternehmen erkennen zunehmend die Chancen, die KI für so gut wie alle Aspekte ihrer Tätigkeit bietet.

Vom Kunden-Engagement, Support und Vertrieb bis hin zu IT-Infrastruktur, Prozessen, Code-Entwicklung und der Bereitstellung von Lösungen findet KI in zahlreichen Branchen immer häufiger Verwendung und gewinnt an Dynamik.

Laut IDC wird der Markt bis 2027 voraussichtlich 423 Milliarden USD übersteigen, mit einer jährlichen Wachstumsrate (CAGR) von 26,9 %, wobei viele Unternehmen ihre KI-Initiativen auf die Verbesserung der operativen Effizienz, des Kundenerlebnisses und der Produktivität konzentrieren.¹

Inmitten dieser rasanten Entwicklung stehen Führungskräfte unter Druck, KI-Lösungen zu identifizieren, auszuwählen, zu entwickeln und bereitzustellen, die ihren Unternehmen einen Wettbewerbsvorteil verschaffen. Das Tempo der KI-Innovation und die Fähigkeit der meisten Unternehmen, ihren KI-Reifegrad zu erhöhen, sind jedoch verschieden. Dies macht es schwierig, den vollen Wert der KI zu erschließen, und führt in vielen Fällen zu mehr Fragen als Antworten.



Unabhängig davon, ob Sie gerade erst mit der KI-Einführung beginnen, mehr über die Auswirkungen von KI auf Ihr Unternehmen erfahren möchten oder herausfinden wollen, wie Sie bestehende KI-Implementierungen skalieren können – dieses E-Book beantwortet viele der aktuellen Fragen zu KI.

¹ IDC FutureScape Webcast: „[Worldwide Artificial Intelligence and Automation 2024 Predictions](#)“. Dokument #US51901124, März 2024.

Welche Arten von KI

Zur bestmöglichen Nutzung von KI sollten Sie möglichst viel über sie wissen, einschließlich der beiden wichtigsten Arten von KI, die heute von Unternehmen eingesetzt werden.

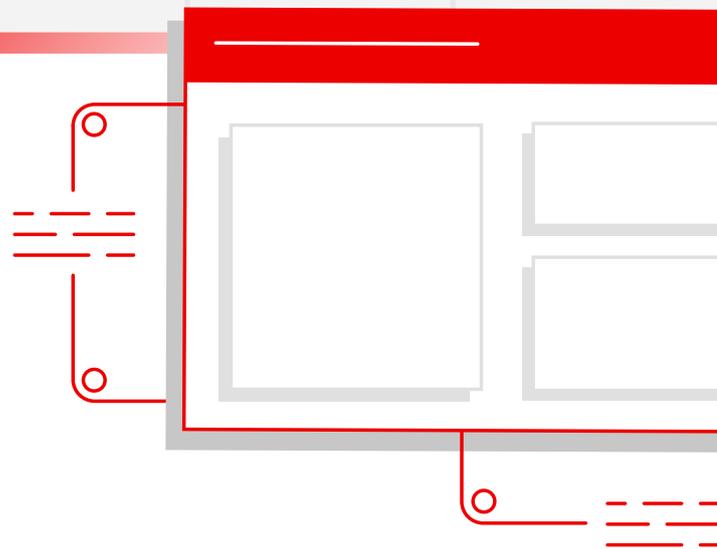
Prädiktive KI: Anhand historischer Daten hilft prädiktive KI Unternehmen, Muster zu erkennen und fundierte Entscheidungen für die Zukunft zu treffen. Prädiktive Modelle ermöglichen Anwendungen wie Bedarfsprognosen, prädiktive Wartung und operative Planung. Die prädiktive KI basiert auf bewährten Techniken der Datenwissenschaft und des maschinellen Lernens (ML), so dass sich die KI durch die Verarbeitung von mehr Daten verbessern kann.



Generative KI (gen KI): Generative KI, die von Deep Learning-Modellen wie Transformers unterstützt wird, kann neue Inhalte wie Text, Bilder und Code erstellen. Sie ist besonders nützlich für Anwendungen wie Chatbots, die automatische Erstellung von Inhalten und kreative Tools. Modelle wie GPTs (Generative Pretrained Transformers) haben die Verarbeitung natürlicher Sprache und kreative Bereiche revolutioniert, indem sie wie von Menschen geschaffene Texte und Bilder erzeugen.

Welche Vorteile bietet die Implementierung von KI?

Das volle Potenzial von KI muss erst noch entdeckt werden. Wenn Sie jedoch verstehen, wie diese sich schnell entwickelnde Technologie Unternehmen unterschiedlicher Größe in vielen Branchen bereits jetzt zugutekommt, können Sie entscheiden, wo in Ihrem Unternehmen KI integriert werden sollte.



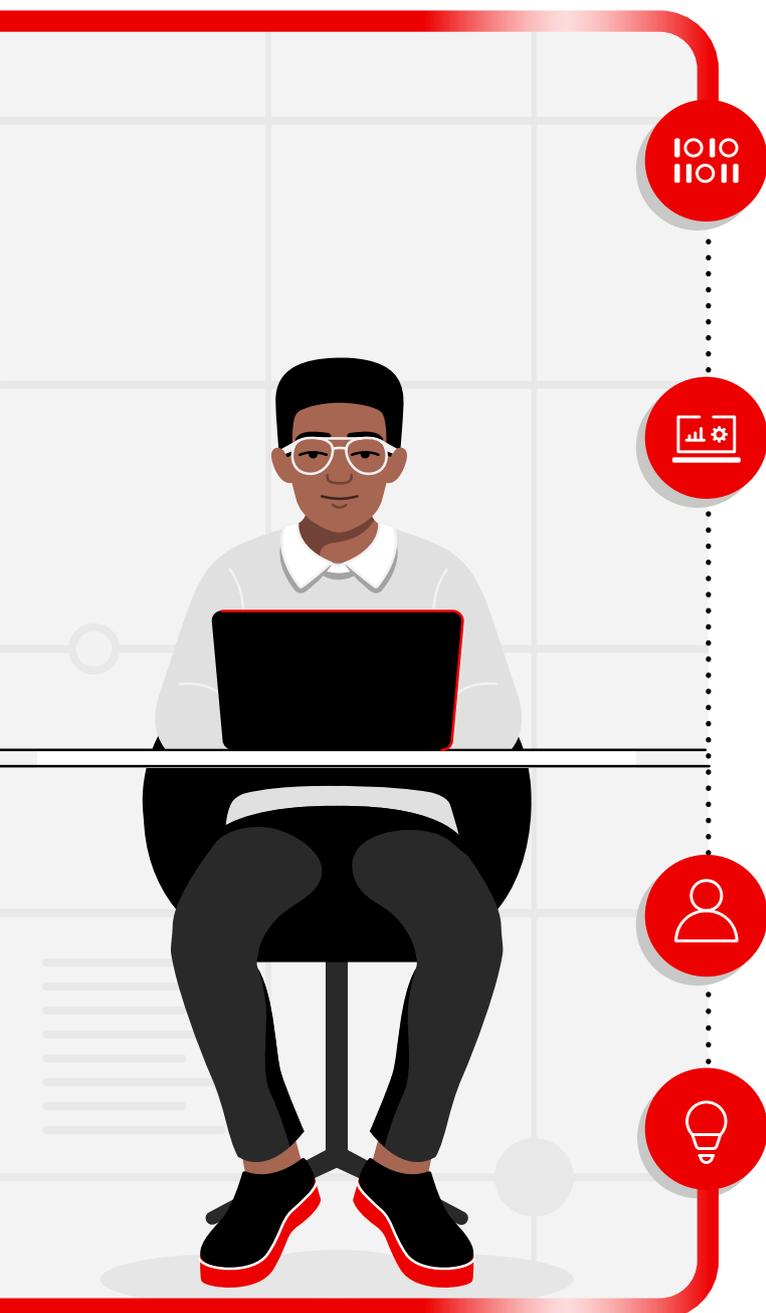
Vorteile von KI und wie diese Ihrem Unternehmen helfen könnten:

Datenvolumen: Angesichts des enormen Datenwachstums haben Unternehmen oft Schwierigkeiten, die riesigen Datenmengen zu verwalten und Insights aus ihnen zu gewinnen. KI kann große Datenmengen schnell verarbeiten und analysieren und dabei wertvolle Insights und Trends aufdecken, die manuell nur schwer zu ermitteln wären.

Operative Ineffizienz: Viele Unternehmen sind sich bewusst, dass ineffiziente Prozesse und Engpässe die Produktivität behindern können, sodass mehr Zeit und Mühe erforderlich sind, um diese Hürden zu überwinden. Die KI-gestützte Automatisierung trägt zur Optimierung der Abläufe bei, wodurch Fehler reduziert und die Prozesseffizienz verbessert werden. Dazu gehören beispielsweise einfache Anwendungen wie das automatische Erstellen von Besprechungsnotizen mit Aktionspunkten und klaren nächsten Schritten oder das schnellere Erstellen von Grafiken und Videos für Websites oder soziale Medien.

Kundenerwartungen: Kunden erwarten personalisierte und reibungslose Erlebnisse. Durch die Analyse von Kundendaten und die Bereitstellung individueller Empfehlungen und kundenspezifischer Interaktionen kann KI den Kundenservice und die Personalisierung verbessern.

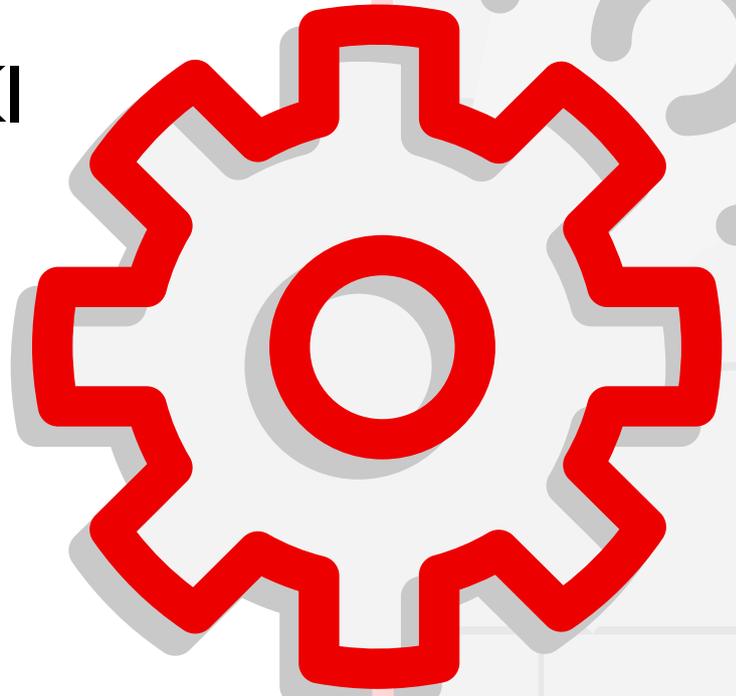
Wettbewerbsfähigkeit auf dem Markt: Um auf einem sich rasch entwickelnden Markt wettbewerbsfähig zu bleiben, ist ständige Innovation erforderlich. KI kann Unternehmen helfen, sich schnell an Marktveränderungen anzupassen und sich so Wettbewerbsvorteile zu sichern. Gen KI kann sogar zum Verfeinern Ihrer Konzepte eingesetzt werden, wenn sie als Ideengeber für Führungskräfte oder bei der Vorbereitung auf wichtige Besprechungen eingesetzt wird.



Der Aufstieg der KI

KI wird seit Jahrzehnten weiterentwickelt und hat Fortschritte in Branchen wie dem Gesundheitswesen, Finanzwesen und der Fertigung ermöglicht.

Der jüngste Aufstieg der gen KI hat jedoch aufgrund ihrer Fähigkeit, menschenähnlich Text, realistische Bilder und sogar Softwarecode zu erstellen, Aufmerksamkeit erregt. Im Gegensatz zu herkömmlicher KI, die Aufgaben automatisiert oder Daten analysiert, ebnet gen KI den Weg für kreative Problemlösungen und die fortschrittliche Erstellung von Inhalten.



Arten von KI-Modellen, die Innovationen beschleunigen

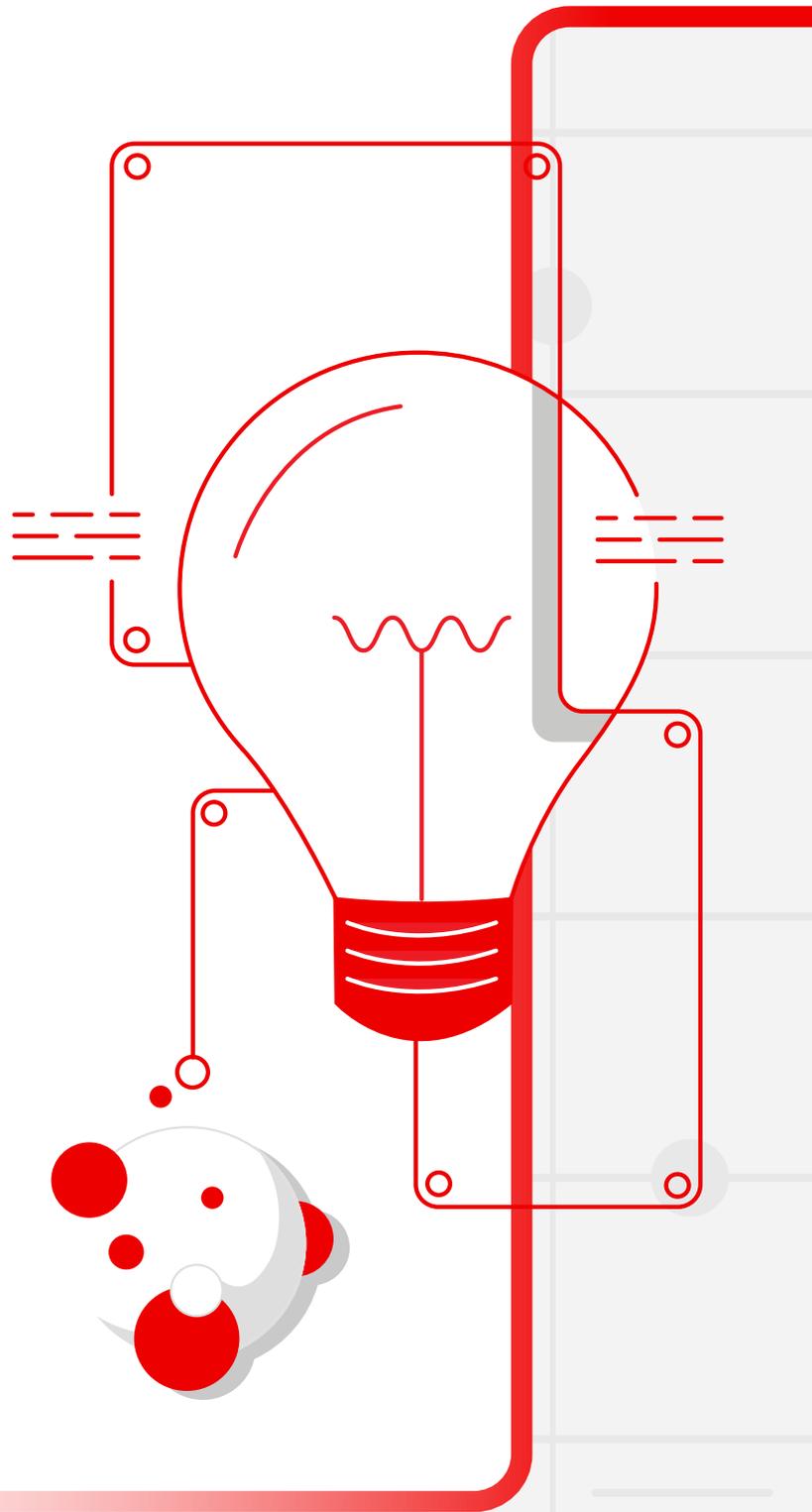
[LLMs \(Large Language Models\)](#) und stabile Diffusionsmodelle gehören zu den KI-Modellen, die für das explosive Wachstum der gen KI verantwortlich sind. LLMs wie GPT sind auf riesigen Datensätzen vortrainiert und können natürliche Sprache verstehen und generieren, was sie für die Automatisierung des Kundensupports, die Erstellung von Marketingtexten und vieles mehr von unschätzbarem Wert macht. Stabile Diffusionsmodelle hingegen schaffen hyperrealistische Bilder, die Innovationen in der Unterhaltung, im Marketing und weiteren Bereichen voranbringen.

Aufkommende Trends, die zu beachten sind

Unternehmen interessieren sich zunehmend für multimodale KI, die Funktionen zur Text-, Bild- und Datenverarbeitung in einem einzigen Modell vereint und vielseitigere Lösungen bietet. Diesen Trends immer einen Schritt voraus zu sein, ist der Schlüssel zur Nutzung des vollen Potenzials von KI im Unternehmensumfeld.

Open Source: Eine Basis für KI-Innovation

Die [KI-Strategie von Red Hat](#) ist tief in Open Source verwurzelt und hilft Unternehmen, gen KI mit Transparenz, Vertrauen und geringeren Kosten voranzutreiben. Durch den Einsatz der Open [Hybrid Cloud](#)-Plattformen von Red Hat können Unternehmen ungehindert innovieren und gleichzeitig die Kontrolle über ihre KI-Lösungen behalten.



[Erfahren Sie mehr über LLMs und deren Funktionsweise](#)

Kontrolle über LLMs mit Open Source

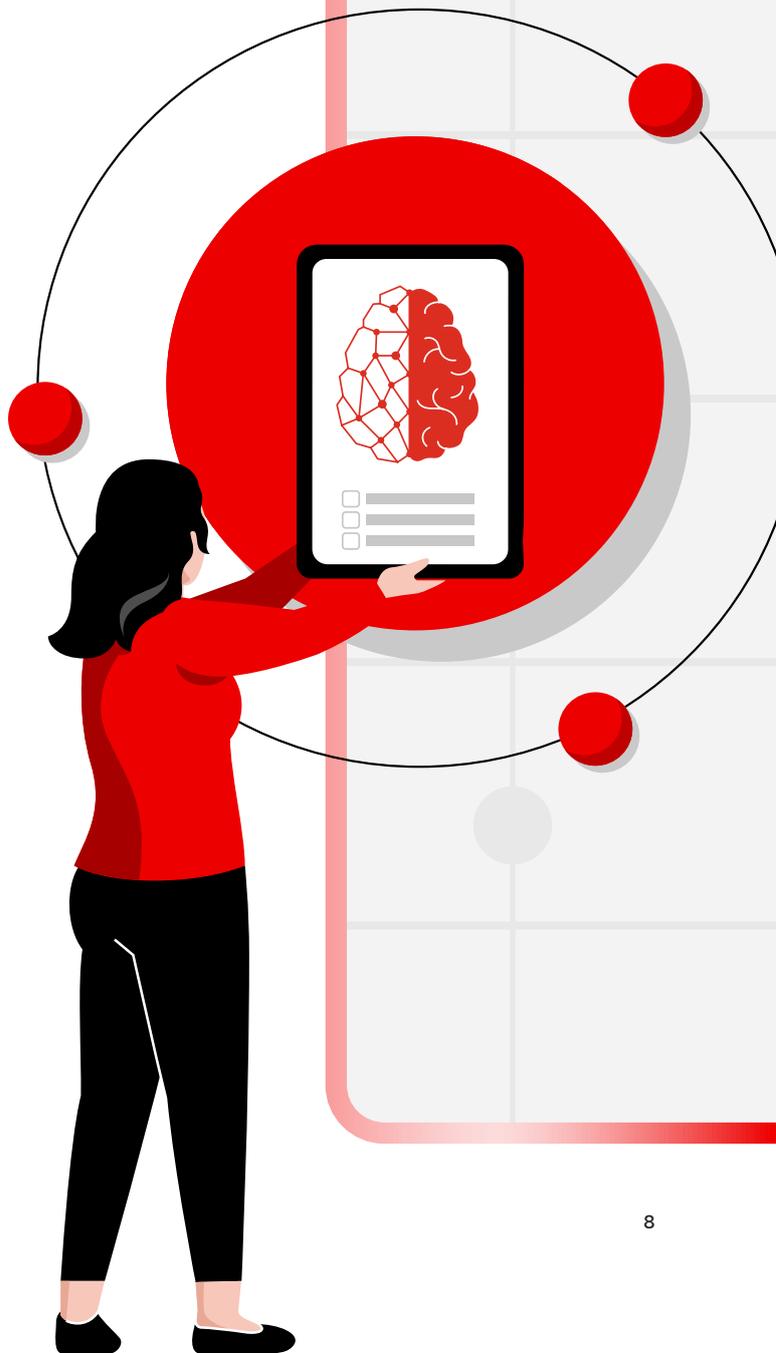
Während gen KI fast sämtliche Aspekte des Geschäftslebens verändert – von der Herstellung von Software bis hin zur Art und Weise, wie wir kommunizieren – ist es nicht ungewöhnlich, dass die als Teil einer gen KI-Funktionalität verwendeten Modelle (LLMs und andere Arten) vom Serviceanbieter streng kontrolliert werden. Das bedeutet, dass es für ein Unternehmen nicht einfach ist, die Fähigkeiten eines gen KI-Dienstes ohne spezielle Kompetenzen und manchmal hohen Aufwand (sowohl finanziell als auch zeitlich) zu testen.

Ohne Einblick in die Datensätze, aus denen das Modell erstellt wurde, oder ohne Details zur Nutzung der Daten durch das Modell sind Unternehmen potenziellen Risiken in Bezug auf KI-generierte Inhalte ausgesetzt. Was passiert, wenn ein Modell zur Codegenerierung mit urheberrechtlich geschütztem Quellcode trainiert wurde? Ist jeglicher von diesem Modell erzeugter Code ebenfalls Teil des urheberrechtlich geschützten Codes? Viele Fragen wie diese sind noch nicht vollständig geklärt, aber da Unternehmen die möglichen erheblichen Konsequenzen erkennen, wenden sie sich der Open Source-KI zu.

Der KI-Ansatz von Red Hat ist in Open Source verwurzelt, ebenso wie unser Support für Open Source-Modelle wie die **[IBM Granite-Familie](#)** von Basismodellen.

Mit **[InstructLab](#)**, einer communitybasierten Lösung zur Verbesserung der LLM-Funktionen, tragen die KI-Lösungen von Red Hat sogar direkt zur KI-Modellentwicklung bei.

[InstructLab auf GitHub erkunden](#)



Auswählen des richtigen KI-Modells

Verschiedene KI-Modelle können für unterschiedliche Use Cases genutzt werden.

Modelle für prädiktive KI und generative KI könnten zusammen in einer einzigen Anwendung oder einem einzigen Dienst verwendet werden. Jedes dieser Modelle ist mit unterschiedlichen Kosten und besonderen Vorteilen verbunden, kann aber Unternehmen bei der Verkürzung der Markteinführungszeit für den ersten Proof of Concept unterstützen. Image-Segmentierung, Voice to Text und Bilderkennungsmodelle sind gängige und sehr leistungsfähige Beispiele. Wichtig ist jedoch die Bewertung der Eignung für Ihren Use Case.

Die auf großen Datenmengen trainierten Basismodelle bieten eine hohe Flexibilität ihrer Fähigkeiten. Allerdings kann ihr großer Umfang die Kosten in die Höhe treiben, den Verwaltungsaufwand erhöhen und die Komplexität steigern, was bedeutet, dass sie möglicherweise nicht für alle Use Cases geeignet sind.

Kleinere, feinabgestimmte Modelle (die immer noch Teil der KI-Familie sind) können eine besser geeignete Lösung darstellen, wenn ein Tuning eines bestehenden Modells auf Ihre Anforderungen bevorzugt wird. Wenn Sie sich für ein Modell entscheiden, können Sie eine vorgefertigte Option wählen, die leicht verfügbar ist und einfach in Ihre Systeme integriert werden kann. Ein gängiges Beispiel für diese Art Modell ist ein LLM, ein leistungsfähiges Tool, das bereits mit großen Datenmengen trainiert wurde.

Haben Sie spezielle geschäftliche Anforderungen, Bedenken hinsichtlich des Datenschutzes oder den Wunsch nach mehr Kontrolle über das Verhalten des Modells, kann es erforderlich sein, ein benutzerdefiniertes Modell zu entwickeln und selbst zu hosten.

Modellentwicklung im Vergleich zu Modell-Tuning

Die Entwicklung eines KI-Modells von Grund auf kann ein umfangreiches Projekt sein. Sie müssen große Datensätze sammeln und vorbereiten, die für die geschäftlichen Use Cases Ihres Unternehmens relevant sind. Dann müssen Sie einen geeigneten Algorithmus auswählen und ihn mit Ihren Daten trainieren. Dieser Prozess erfordert erhebliche Rechenleistung und Fachwissen, was ihn zeit- und ressourcenaufwendig macht. Die Entwicklung eines traditionellen oder Basismodells kann zwar eine benutzerdefinierte Lösung bieten, ist aber nicht immer der effizienteste Weg.

Bei der Abstimmung eines Basismodells geht es dagegen um die Anpassung eines vortrainierten Modells an Ihre spezifischen Anforderungen. Ein gängiger Ansatz ist das Transfer-Lernen, bei dem ein auf einem großen Datensatz trainiertes Modell auf einen kleineren, domainspezifischen Datensatz umtrainiert wird. Dies ermöglicht es dem Modell, das während des anfänglichen Trainings erlernte, allgemeine Wissen zu behalten, während es sich an die Nuancen Ihrer spezifischen Daten anpasst.

Fine Tuning Ihres Modells

Ein anderer Ansatz ist das Fine Tuning, bei dem Sie die Parameter des Modells anpassen, um die Performance für Ihre spezielle Aufgabe zu verbessern. Die Modellparameter beziehen sich auf die Variablen eines ausgewählten Modells, die durch Anpassung der gegebenen Daten an das Modell geschätzt werden können. Beim Fine Tuning ist es möglicherweise erforderlich, die Lernrate zu ändern, die Architektur des Modells zu modifizieren oder bestimmte Schichten des Modells intensiver zu trainieren als andere. Diese Techniken tragen dazu bei, das Wissen des Modells zu erweitern und es für Ihren speziellen Use Case effektiver zu machen.

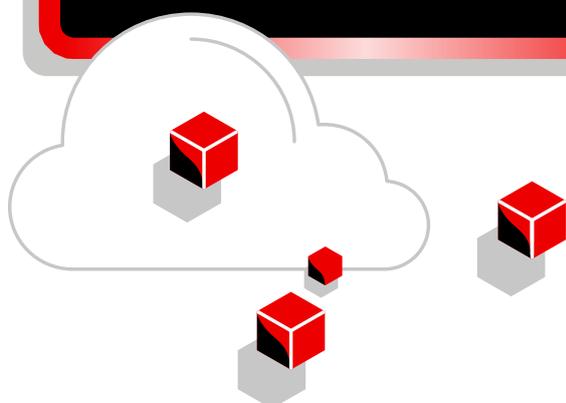


InstructLab verfolgt einen Fine Tuning-Ansatz mit dem Ziel, das erforderliche KI-Fachwissen zu reduzieren und das Hinzufügen unternehmenseigener Kenntnisse zu bestehenden Granite-Basismodellen zu vereinfachen.

Alternativen zum Fine Tuning der Modelle

Forschungsteams suchen nach Möglichkeiten, Basismodelle für mehr Geschwindigkeit und Effizienz zu tunen. Zu den gängigen Verfahren zählt RAG (Retrieval-Augmented Generation). Dabei handelt es sich um eine Technik zum Abrufen von Fakten aus einer externen Quelle, in der zusätzliches Wissen (Kontext) kodiert wurde.

RAG stützt sich auf die Verwendung einer oder mehrerer externer Datenbanken (Vektordatenbanken), die zusätzlichen Kontext zu der Frage liefern, die dem gen KI-Modell gestellt wird. Ein weiterer neuer Ansatz sind agentische KI-Systeme, bei denen eine Reihe von gen KI-Agenten zusammenarbeiten, um externe Systeme nach Wissen abzufragen – etwa interne Datenbanken, Firmen-Intranets oder das Internet – und die gen KI-Modelle mit den genauesten und aktuellsten Informationen zu versorgen.



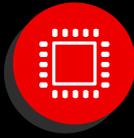
Das letzte Beispiel ist das Prompt Tuning, bei dem KI-Modelle Hinweise oder Frontend Prompts erhalten, einschließlich zusätzlicher Wörter oder KI-generierter Zahlen, die die Modelle zu einer gewünschten Entscheidung führen. Das Ergebnis einer RAG-Abfrage würde einen zusätzlichen Kontext für einen Prompt bilden, wobei Prompt Tuning und RAG zusammenwirken. Die Kombination aus Fine Tuning eines bestehenden LLM und dem Einsatz von RAG-Techniken und Prompt Tuning ermöglicht es Organisationen mit begrenzten Daten, ein Basismodell auf eine spezielle Aufgabe abzustimmen.

Die Infrastruktur, die Ihr KI-Modell unterstützt, ist ebenso wichtig wie das Modell selbst. Verschiedene Aufgaben erfordern verschiedene Arten von Hardware.



Central Processing Unit (CPU)

Herkömmliche Prozessoren, die allgemeine Berechnungen durchführen. Sie sind vielseitig, aber können möglicherweise nicht effizient mit umfangreichen KI-Workloads umgehen.



Graphics Processing Unit (GPU)

Spezialisierte Prozessoren, die für die parallele Verarbeitung von Aufgaben ausgelegt sind und sich daher ideal für das Training von Deep Learning-Modellen eignen, wobei große Datenmengen gleichzeitig verarbeitet werden müssen.



Neural Processing Unit (NPU)

Ein neuartiger Prozessortyp, der speziell für KI-Aufgaben entwickelt wurde und für bestimmte Modelle eine noch höhere Effizienz und Geschwindigkeit bietet.

Die Rolle der Hybrid Cloud bei der Einführung von KI in Unternehmen

Hybrid Cloud-Umgebungen spielen eine wichtige Rolle bei der Einführung von KI. Eine Hybrid Cloud kombiniert die Infrastruktur vor Ort mit Public- und Private Cloud-Ressourcen und bietet Flexibilität bei der Art und Weise, wie und wo Sie KI-Workloads einsetzen und verwalten. So können Sie beispielsweise Ihre KI-Modelle mit leistungsstarken, cloudbasierten GPUs trainieren und sie dann aus Gründen der Sicherheit oder Compliance lokal oder in einer Private Cloud bereitstellen. Ein wichtiger Aspekt bei der Nutzung des Hybrid Cloud-Konzepts ist daher die Konsistenz der Tools und der gewählten Plattform.

Der Open Hybrid Cloud-Ansatz von Red Hat hilft Unternehmen bei der Integration von KI über verschiedene Umgebungen hinweg und verbessert die Konsistenz, Skalierbarkeit und Flexibilität. Mit diesem Ansatz können Sie Ihre KI-Workloads über mehrere Cloud-Umgebungen hinweg, On-Premise oder am Netzwerkrand verwalten, die Datenplatzierung optimieren und eine reibungslose Datenmigration fördern, was die Einführung von KI auf Unternehmensebene erleichtert.

Wenn Sie die KI-Modelle, die Daten und die Infrastruktur verstehen, können Sie die Komplexität der KI-Einführung besser bewältigen und ihr volles Potenzial nutzen.

Was Sie für den Einstieg benötigen

Wie bei der Einführung anderer, neuer Technologien gibt es auch bei der KI Hürden, die ein Unternehmen überwinden muss, um erfolgreich zu sein.

Bewerten Sie anhand der folgenden Überlegungen die Bereitschaft Ihres Unternehmens und ermitteln Sie die Bereiche, auf die Sie sich zur Beschleunigung Ihrer KI-Einführung konzentrieren müssen.



Prüfen der Qualität und Verfügbarkeit von Daten:

Der Zugang zu hochwertigen, relevanten Daten ist für die KI unerlässlich. Die Datenqualität ist entscheidend für das Training präziser KI-Modelle. Daher ist die Bewertung der Vollständigkeit, Genauigkeit und Relevanz Ihrer Daten wichtig.



Bewerten der technologischen Infrastruktur: Ermitteln Sie, ob Ihre aktuelle Infrastruktur KI-Workloads bewältigen kann. Dazu gehört auch die Bewertung der Verfügbarkeit leistungsstarker Rechenressourcen, Storage-Lösungen, Automatisierung und Netzwerkkapazitäten.



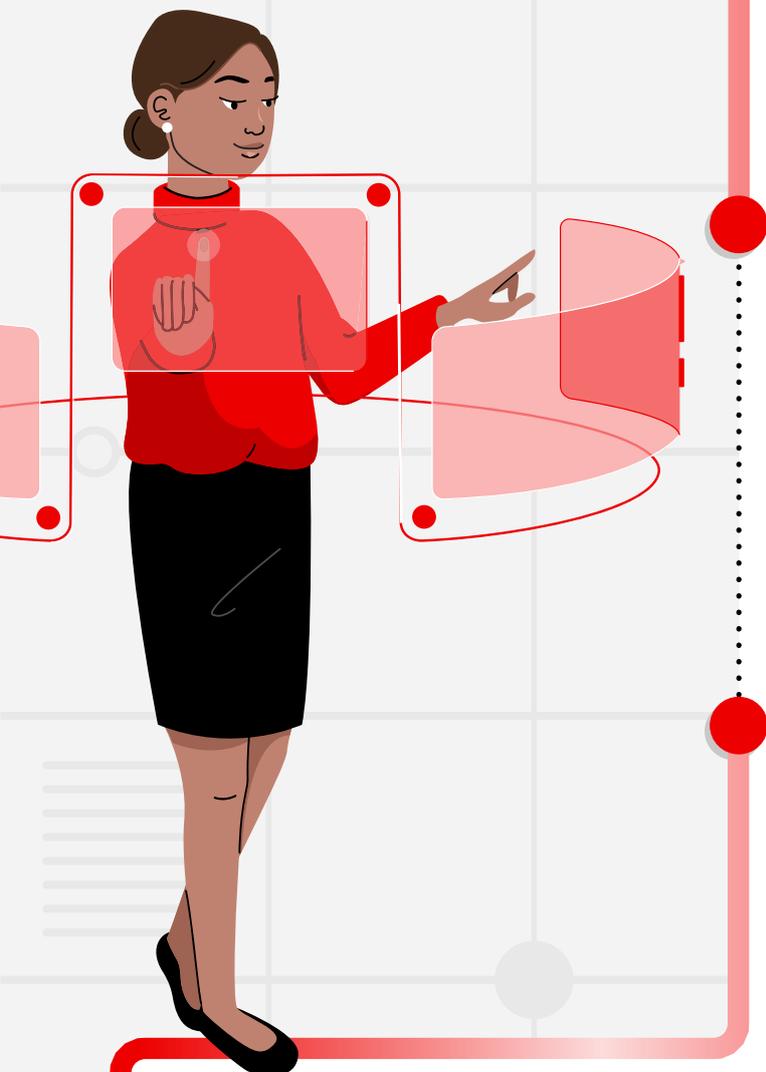
Ermitteln des Bedarfs an Kompetenzen: Bewerten Sie die Verfügbarkeit von Fachwissen zu KI in Ihrem Unternehmen. Dies beinhaltet die Bewertung der aktuellen vorhandenen Kompetenzen und die Ermittlung des Bedarfs an Training oder speziellen Kompetenzen.



Prüfen der strategischen Ausrichtung: Stellen Sie sicher, dass Ihre KI-Initiativen mit den Zielen und Strategien Ihres Unternehmens übereinstimmen. KI-Projekte sollten die übergeordneten strategischen Ziele Ihres Unternehmens unterstützen und einen messbaren Geschäftswert liefern.

Der Einstieg mit KI

Geschwindigkeit und Ausmaß der Einführung von KI in einem Unternehmen hängen von vielen Faktoren ab. Bei fast allen technologischen Modernisierungsprojekten ist es jedoch sinnvoll, langsam zu starten und schrittweise zu wachsen.



Die folgenden 8 Schritte helfen Ihrem Unternehmen, erste Schritte mit KI zu unternehmen und ihre Einführung voranzutreiben:

1

Prüfen von Fähigkeiten und Zielen

Beginnen Sie mit einer Bewertung der aktuellen Fähigkeiten, der Infrastruktur und der strategischen Ziele Ihres Unternehmens. Bestimmen Sie, ob KI mit Ihren generellen Zielen übereinstimmt, und ermitteln Sie die potenziellen Bereiche, in denen KI einen Mehrwert schaffen kann. Diese erste Bewertung wird Ihnen helfen, eine klare Richtung für die Einführung von KI vorzugeben.

2

Identifizieren von Use Cases und KI-Teams

Prüfen Sie die Möglichkeiten, die KI in Ihrem Unternehmen bieten kann. Bilden Sie ein spezielles KI-Anwendungsteam mit funktionsübergreifenden Mitgliedern wie Entwicklerinnen und Entwicklern, Domain-Fachleuten, Data Scientists und IT-Spezialistinnen und -Spezialisten zur Leitung der Initiative. Ein klar definierter Use Case lenkt Ihre Maßnahmen zur Einführung von KI und bündelt Ihre Ressourcen.

3

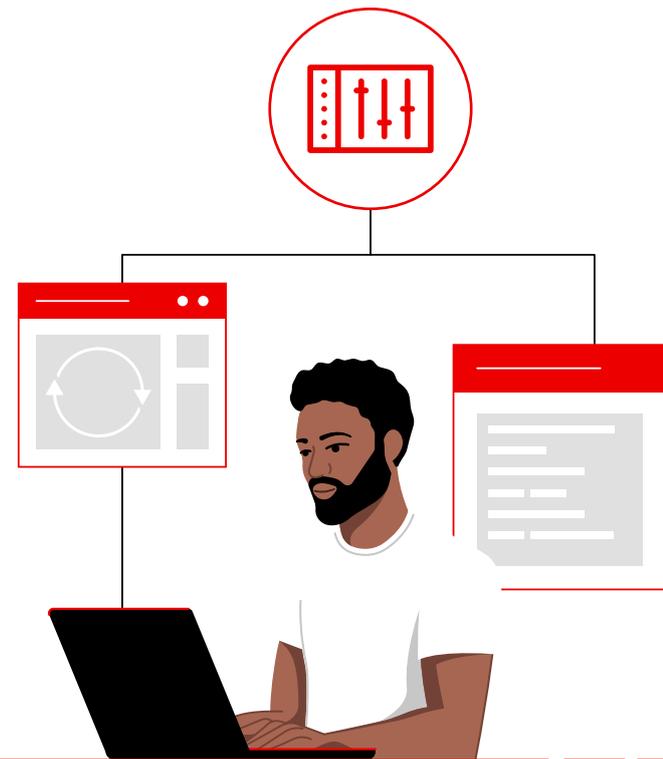
Modellauswahl

Wählen Sie das geeignete KI-Modell auf der Grundlage Ihres konkreten Use Cases. Unabhängig davon, ob es sich um ein LLM für gen KI oder ein prädiktives Modell für die Datenanalyse handelt – stellen Sie sicher, dass die Fähigkeiten des Modells den Zielen Ihres Use Cases gerecht werden. Berücksichtigen Sie Faktoren wie die Komplexität des Modells, seine Skalierbarkeit und die Kompatibilität mit Ihren bestehenden Systemen.

4

Test- und Validierungsschleifen

Definieren Sie klare Erfolgskriterien für Ihre KI-Implementierung, wie Performancekennzahlen, Genauigkeitsraten oder Geschäftsziele. Führen Sie Test- und Validierungsschleifen zur kontinuierlichen Bewertung der Wirksamkeit Ihres Modells ein. Regelmäßiges Feedback aus diesen Schleifen hilft beim Fine Tuning des Modells und hält Ihre KI-Einführung auf Kurs.



5

Modell-Tuning

Passen Sie das ausgewählte Modell anhand der Daten Ihres Unternehmens an. Bei diesem Anpassungsprozess wird das Modell mit relevanten Daten versorgt, um seine Genauigkeit und Relevanz für spezifische Use Cases in Ihrem Unternehmen zu verbessern. Mittels Fine Tuning wird sichergestellt, dass sich das Modell an den besonderen Kontext und die Anforderungen Ihres Unternehmens anpasst.

6

Training mit synthetischen Daten

Erwägen Sie, Ihre KI-Modelle mit synthetischen Daten weiter zu trainieren und zu verbessern. Dieser Ansatz, bei dem Methoden wie LLM-Lehrer und LLM-Schüler zum Einsatz kommen, ermöglicht Ihnen die Generierung hochwertiger Trainingsdaten, wenn reale Daten fehlen oder vertraulich sind. Synthetische Daten können zur Verbesserung der Robustheit und Performance des Modells beitragen, ohne die Privatsphäre zu gefährden.

7

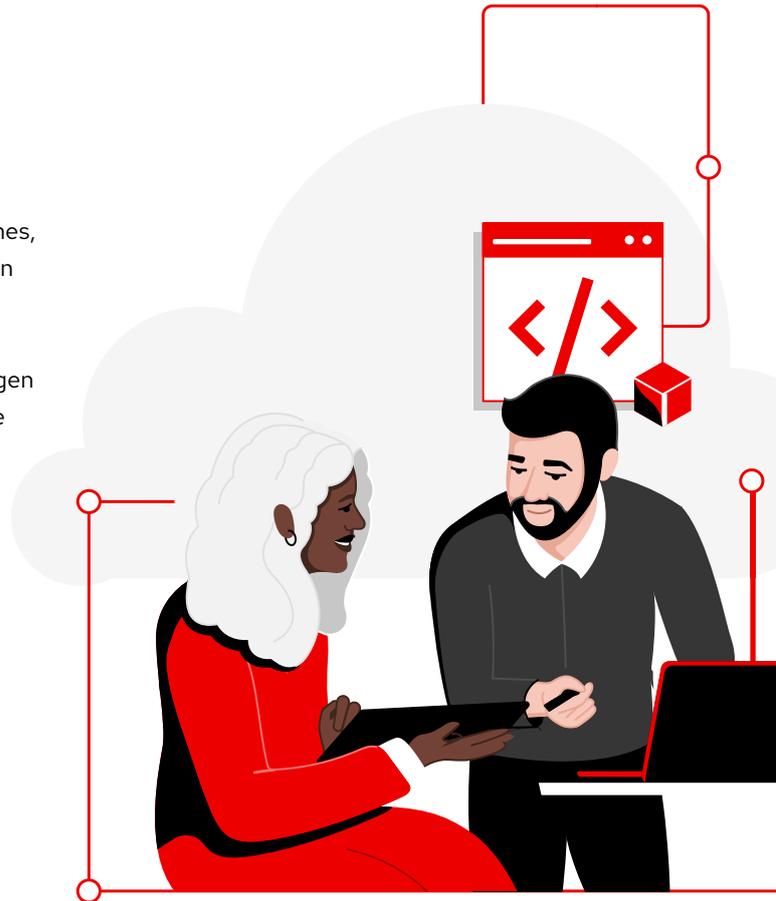
Überwachung von Abweichungen

Die Überwachung von Abweichungen bietet ein allgemeines, inhaltsbezogenes Monitoring. Anstelle eines strukturierten Konfigurationsmanagements verfolgt die Überwachung von Abweichungen Änderungen am Inhalt des lokalen Dateisystems. Sie kann Ihnen helfen, etwaige Abweichungen oder Verschlechterungen der Modellgenauigkeit im Laufe der Zeit zu erkennen und zu beheben. Die kontinuierliche Überwachung stellt sicher, dass das Modell wirksam und relevant bleibt, wenn sich Bedingungen ändern.

8

Hilfe von Fachleuten

Wenn Ihr internes KI-Fachwissen noch im Aufbau begriffen ist, kann die Hinzuziehung externer Fachleute wie Red Hat® Consulting von großem Vorteil sein. Die Fachleute von Red Hat können Sie durch die komplexen Aspekte der KI-Einführung leiten, wertvolle Insights liefern und Ihnen Zugang zu Training verschaffen. Red Hat Consulting kann Ihre KI-Einführung beschleunigen und die Erfolgswahrscheinlichkeit erhöhen.



Die KI-Einführung hängt von Zusammenarbeit ab

Der Aufbau eines Teams mit einer Vielzahl von Fähigkeiten ist für den Erfolg von gen KI-Projekten entscheidend.²

- **Business Leader** vertreten die Menschen, die diese Lösung nutzen oder von ihr betroffen sind.
- **KI-Fachleute** tunen, warten und aktualisieren gen KI-Modelle.
- **Data Scientists** bereiten die Daten vor und stellen korrekte, unverzerrte Trainingsdaten für Modelle bereit.
- **Beauftragte für Ethik und Compliance** stellen sicher, dass gen KI-Initiativen den Vorschriften entsprechen.
- **IT Operations Specialists** integrieren Lösungen in die bestehende Infrastruktur und setzen Sicherheitsrichtlinien durch.
- **Entwicklungsteams und -Communities** müssen von Anfang an einbezogen werden, um Open Source-Tools, Frameworks und Best Practices für die Einführung von KI gemeinsam zu entwickeln, zu nutzen und zu verbessern. So wird sichergestellt, dass der Einsatz von KI mit einem geschäftlichen Nutzen verbunden ist.

² Kearney: „[Standing up tiger teams to tackle generative AI complexity](#)“, 15. November 2023.

Einführen und Skalieren mit Red Hat

Red Hat AI bietet Vertrauen, Auswahl und Konsistenz in der Hybrid Cloud, um die Einführung von KI in Unternehmen zu beschleunigen.

Das Portfolio von Red Hat AI umfasst Red Hat Enterprise Linux® AI für einzelne Linux-Serverumgebungen und Red Hat OpenShift® AI für verteilte Kubernetes-Plattformumgebungen sowie integrierte MLOps-Funktionen (Machine Learning Operations). Beide Lösungen basieren auf Open Source-Technologien und -Modellen. Sie helfen Unternehmen, bei KI-Innovation führend zu bleiben, das Forschungstempo zu beschleunigen und den Zugang zu modernsten Tools und Technologien zu demokratisieren.

Das umfangreiche Partnernetzwerk von Red Hat kann Ihre KI-Fähigkeiten weiter stärken. So setzt NVIDIA, das für die Etablierung von GPUs bekannte, bedeutende KI-Unternehmen seine Partnerschaft mit Red Hat mit dem Ziel fort, die Leistungsfähigkeit von KI durch die Bereitstellung einer für KI-Workloads optimierten End-to-End-Plattform für Unternehmen zu erschließen. NVIDIA unterstützt Unternehmen bei der Einführung von GPU-beschleunigtem Computing für KI- und High Performance Computing-Anwendungen.

„Red Hat und NVIDIA blicken auf eine lange Geschichte enger Zusammenarbeit zurück, und Red Hat Enterprise Linux AI demonstriert unseren gemeinsamen Fokus darauf, Entwicklungs- und Forschungsteams, die die nächste Generation von KI-Technologien und -Anwendungen entwickeln, umfassende Rechenleistung und Software zur Verfügung zu stellen.“³

Justin Boitano, Vice President, Enterprise Products, NVIDIA

³ Red Hat Pressemitteilung: „[Red Hat Delivers Accessible, Open Source Generative AI Innovation with Red Hat Enterprise Linux AI](#)“, 7. Mai 2024.



Ein genauerer Blick auf Red Hat Enterprise Linux AI

Red Hat Enterprise Linux AI besteht aus 4 verschiedenen Basiskomponenten:

1

Open Source Granite-Modelle

Red Hat Enterprise Linux AI enthält komplett durch Red Hat unterstützte Open Source Granite-Modelle. Diese flexiblen Modelle ermöglichen Ihnen die Erstellung benutzerdefinierter Sprachmodelle und deren öffentliche oder private Nutzung.

2

Modellanpassung mit InstructLab

InstructLab ist ein Open Source-Projekt unter der Leitung von Red Hat und IBM. Es passt KI-Modelle mit spezifischem Wissen an und erzeugt synthetische Trainingsdaten. Da es sich um ein Befehlszeilen-Tool handelt, das sich in ein Git Repository integrieren lässt, können Nutzende unkompliziert Fähigkeiten hinzufügen und Modelle trainieren.

3

Optimiertes, startfähiges Red Hat Enterprise Linux für Granite-Modelle und InstructLab

Granite-Modelle und InstructLab-Tools laufen auf einem speziellen, für KI optimierten Red Hat Enterprise Linux-Image, das mit praktisch jeder Hardware und Cloud-Umgebung kompatibel ist. Dieses Setup ermöglicht eine effiziente Performance mit High End-GPUs, die für schnelles Training und Deployment von Modellen notwendig sind.

4

Support für Unternehmen und Regelung zum geistigen Eigentum

Subskriptionen für Red Hat Enterprise Linux AI umfassen Unternehmens-Support, einen kompletten Produkt-Lifecycle beginnend mit dem Granite 7B-Modell und Software sowie eine Regelung zur Haftungsfreistellung von Red Hat.

Red Hat Enterprise Linux AI hilft, gen KI-Anwendungen zum Leben zu erwecken

Für Unternehmen, die gerade erst in gen KI einsteigen, bietet Red Hat Enterprise Linux AI einsatzbereite LLMs und Code-Sprachmodelle in einer einzigen Server-Entwicklungs- und Inferenzumgebung.

Damit steht eine einheitliche Umgebung mit Modellen und Tools zur Verfügung, die den Einstieg in gen KI und die Anpassung der Modelle an Ihre Geschäftsdaten erleichtert, ohne dass umfangreiches KI-Fachwissen oder eine entsprechende Infrastruktur erforderlich sind.

Mit der Regelung zur Haftungsfreistellung und vollem Support durch Red Hat reduziert Red Hat Enterprise Linux AI Risiken. Außerdem bietet es einen vereinfachten Ansatz für gen KI, der für Entwickler und Domain-Experten leichter zugänglich ist, sodass diese zusammenarbeiten und die Zeit bis zum Erreichen von Geschäftsergebnissen verkürzen können.

Warum Red Hat Enterprise Linux AI?

LLMs für Unternehmen

Open Source-lizenzierte IBM Granite-LLMs sind unter der Apache 2.0-Lizenz enthalten und erhalten eine Regelung zur Haftungsfreistellung sowie vollständigen Support durch Red Hat.

Community-Zusammenarbeit

InstructLab ermöglicht ein vereinfachtes Experimentieren mit gen KI-Modellen sowie optimierte Modellanpassung.

Cloudnative Skalierbarkeit

Mit dem Image-Modus von Red Hat Enterprise Linux können Sie Ihre KI-Plattform als Container Image verwalten und so Ihren Skalierungsansatz optimieren.

Beschleunigung und KI-Tools

Open Source-Hardwarebeschleuniger und optimierte Deep Learning-Features sorgen für schnellere Ergebnisse.

Mit Red Hat OpenShift AI skalieren

Red Hat OpenShift AI bietet eine integrierte MLOps-Plattform für Entwicklung, Training, Tuning, Bereitstellung und Überwachung KI-gestützter Anwendungen sowie von prädiktiven und Basismodellen in großem Umfang in Hybrid Cloud-Umgebungen.

Red Hat OpenShift AI basiert auf Red Hat OpenShift und bietet ein konsistentes, optimiertes und automatisiertes IT-Erlebnis beim Umgang mit Workload- und Performanceanforderungen von KI/ML-Projekten. MLOps-Praktiken können Unternehmen dabei helfen, schnell auf KI-Innovationen zu reagieren und KI-fähige Anwendungen zügiger in die Produktion zu bringen.

Experimentieren Sie in der Red Hat OpenShift AI Sandbox

Mit von Open Data Hub und anderen Open Source-Projekten kuratierten Komponenten bietet Red Hat OpenShift AI Data Scientists und Entwicklungsteams eine leistungsstarke, offene, hybride KI/ML-Plattform zum Erfassen von Insights aus Daten und zum Erstellen von KI-gestützten Anwendungen.

[Testen Sie es in unserer Entwicklungs-Sandbox >](#)

Warum Red Hat OpenShift AI?

Modellbereitstellung skalieren

Es können Modelle für die Integration in intelligente Anwendungen vor Ort, in der Public Cloud oder am Edge bereitgestellt werden. Diese Modelle können basierend auf Änderungen am Quell-Notebook neu erstellt, bereitgestellt und überwacht werden.

Weniger Zeitaufwand für das KI-Infrastrukturmanagement

Geben Sie Ihren Teams On-Demand-Zugriff auf Ressourcen, damit sie sich auf die Datenrecherche und das Entwickeln von Anwendungen konzentrieren können, die Ihr Unternehmen voranbringen.

Getestete und unterstützte KI/ML-Tools

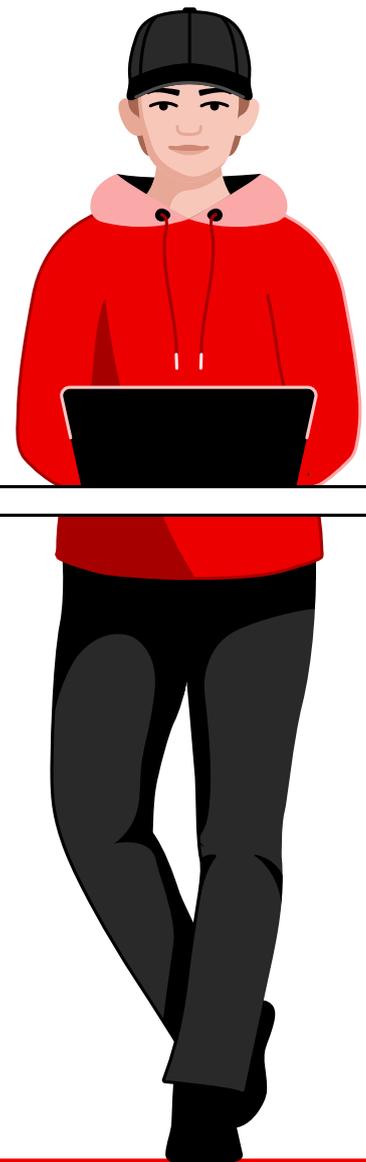
Red Hat verfolgt, integriert, testet und unterstützt bekannte KI/ML-Tools und die Modellbereitstellung auf der Red Hat OpenShift Anwendungsplattform.

Flexibilität in der gesamten Hybrid Cloud

Red Hat OpenShift AI ist wahlweise als selbst gemanagte Software oder als vollständig gemanagter Cloud Service zusätzlich zu Red Hat OpenShift verfügbar. Es bietet eine sicherheitsorientierte und flexible Plattform, bei der Sie selbst entscheiden können, wo Sie Ihre Modelle entwickeln und bereitstellen möchten: On-Premise, in der Public Cloud oder sogar am Netzwerkrand.

Anwendung unserer Best Practices

Red Hat Consulting bietet Services, mit denen Sie Red Hat OpenShift AI in vollem Umfang installieren, konfigurieren und nutzen können. Unabhängig davon, ob Sie ein Red Hat OpenShift AI-Pilotprojekt ausführen möchten oder Unterstützung beim Aufbau Ihrer MLOps-Basis benötigen – Red Hat Consulting bietet Ihnen Support und Mentoring.



Red Hat bietet ein komplettes Technologieportfolio, bewährtes Fachwissen und strategische Partnerschaften, damit Sie Ihre gen KI-Ziele erreichen können. Nutzen Sie eine Basis für das Entwickeln und Bereitstellen von gen KI-Modellen und -Anwendungen sowie Services und Training für eine schnelle Einführung.

Mehr erfahren

Sind Sie bereit für den nächsten Schritt bei der Einführung von KI?

Beschleunigen Sie Ihre KI-Einführung mit der Open Hybrid Cloud-Strategie von Red Hat. Sie verleiht Ihnen die Flexibilität zur Ausführung Ihrer KI-Anwendungen, wo immer Sie diese benötigen.

Beschleunigen Sie Ihre KI/ML-Projekte mit Fachwissen, Consulting und Trainingservices von Red Hat, um Ihrem Unternehmen zu helfen, mit KI seine Ziele zu erreichen.

Mehr über KI/ML-Services erfahren:
red.ht/aiml-consulting

Vereinbaren Sie eine kostenlose Discovery Session:
redhat.com/consulting

Mehr über Red Hat AI erfahren



[Mehr über Red Hat Enterprise Linux AI lesen](#)



[Erfahren Sie, wie Sie mit Red Hat OpenShift AI skalieren](#)

