





# Modernste KI-Modelle in eine souveräne Umgebung integrieren

Mit KI sicher und erfolgreich in die  
Zukunft

19.11.2025

Marius Kiskemper  
Justus Mrosk

**Atos**

# Persönliche Vorstellung



## Marius Kiskemper

- Lead Data Scientist
- Data & AI Business Line
- Paderborn
- B. Sc. Data Science
- M. Sc. Artificial Intelligence
- > 3 Jahre Hands-On Projekterfahrung im Bereich AI
- > 2 Jahre Arbeit mit Red Hat AI, inkl. strategischer Weiterentwicklung des Portfolios

# Persönliche Vorstellung



## Justus Mrosk

- Data Engineer
- Data & AI Business Line
- Essen
- B. Sc. Wirtschaftsinformatik
- > 2 Jahre Erfahrung mit OpenShift und der Bereitstellung von KI-Anwendungen
- > 3 Jahre Erfahrung mit der Entwicklung von KI basierten Anwendungen

# Warum Atos ?

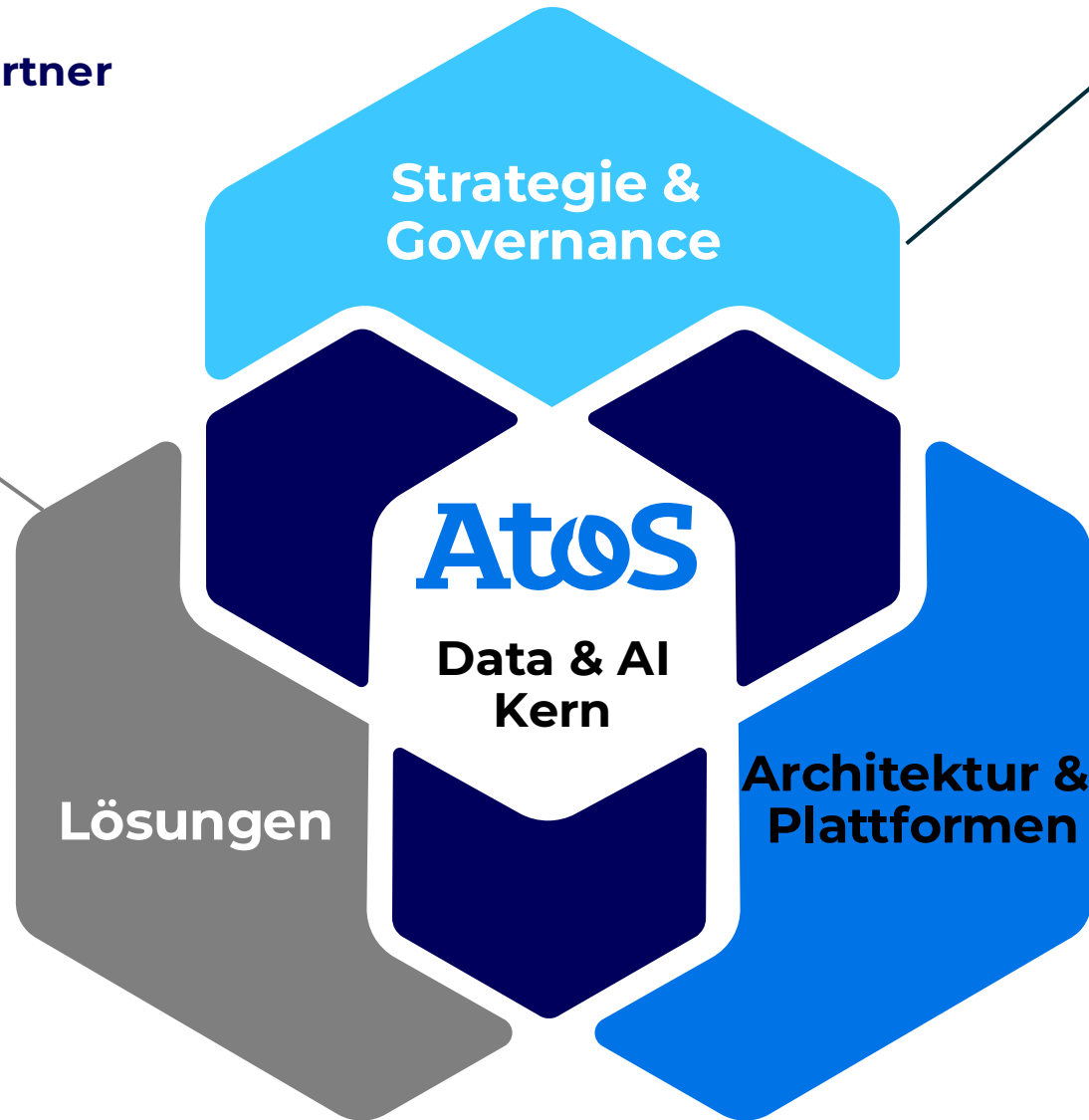
Wir sind der End-to-End-Partner

Daten in umsetzbare  
Erkenntnisse  
verwandeln

Individuell zugeschnittene  
Machine-Learning-  
Anwendungen

Objekterkennung  
Predictive Maintenance  
Geodatenanalyse  
Generative AI-Lösungen

AIRBUS



Datenkompetenz  
stärken

Entwicklung von **KI-Strategien** mit  
**maßgeschneiderten Governance-**  
Rahmenwerken und  
**verantwortungsvoller KI-**  
Anwendung



Aufbau und  
Betrieb robuster  
Datenfundamente

**Cloud-native** Services  
und **MLOps-**  
**Implementierungen**  
auf allen führenden  
Plattformen



OPENSIFT



# Spotlight: Atos als AI-Leader zusammen mit Red Hat

Atos: Experte im Bereich Red Hat AI



# Agenda

- |           |   |           |                    |
|-----------|---|-----------|--------------------|
| <b>01</b> | Die Herausforderung der souveränen Umgebung | <b>05</b> | Modelltraining     |
| <b>02</b> | Deep Dive – Language Models                 | <b>06</b> | Modelldeployment   |
| <b>03</b> | Deep Dive – OpenShift AI Plattform          | <b>07</b> | Modellevaluierung  |
| <b>04</b> | Überblick zur Lösungsarchitektur            | <b>08</b> | Fazit und Ausblick |

# 01

## Die Herausforderung der souveränen Umgebung



# Die Herausforderung der souveränen Umgebung

Kernbestandteile einer sicheren und souveränen Infrastruktur

**Sensible Daten**

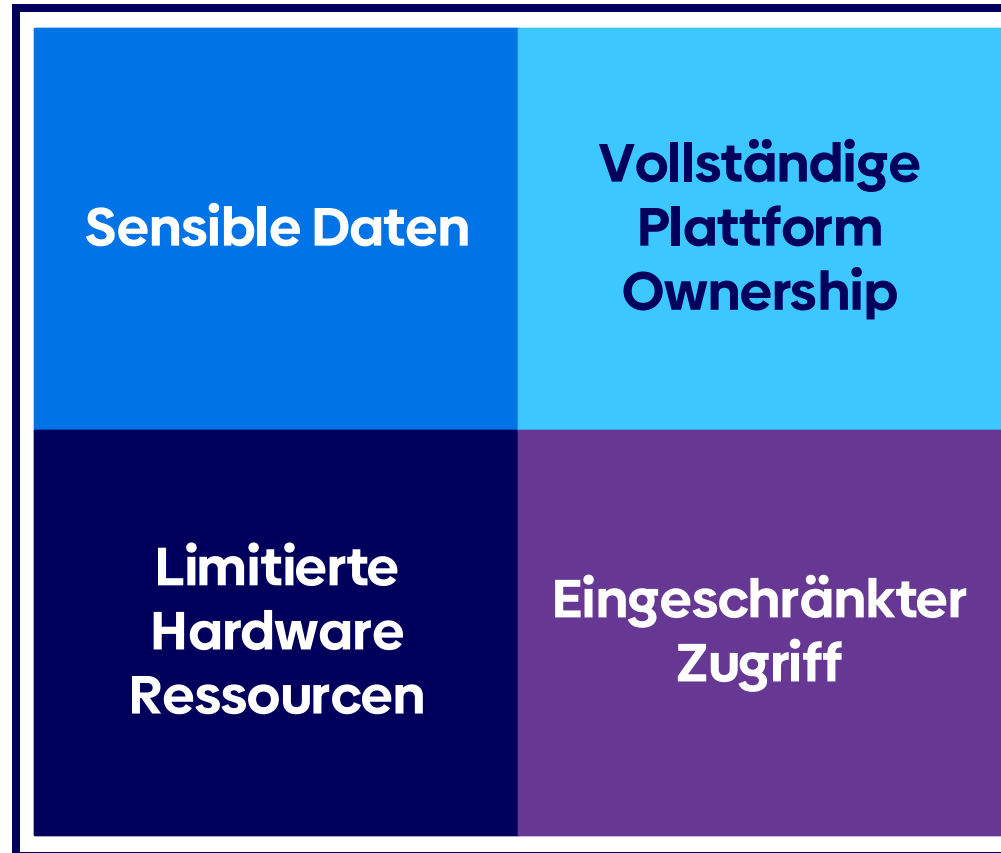
**Vollständige Plattform Ownership**

**Limitierte Hardware Ressourcen**

**Eingeschränkter Zugriff**

# Die Herausforderung der souveränen Umgebung

Kernbestandteile einer sicheren und souveränen Infrastruktur



# Die Herausforderung der souveränen Umgebung

Kernbestandteile einer sicheren und souveränen Infrastruktur



**Sensible Daten**

**Vollständige  
Plattform  
Ownership**

**Limitierte  
Hardware  
Ressourcen**

**Eingeschränkter  
Zugriff**



# Die Herausforderung der souveränen Umgebung

Chancen durch neueste Technologien

**Open Source-Innovationen**

**Einbindung von Künstlicher  
Intelligenz**

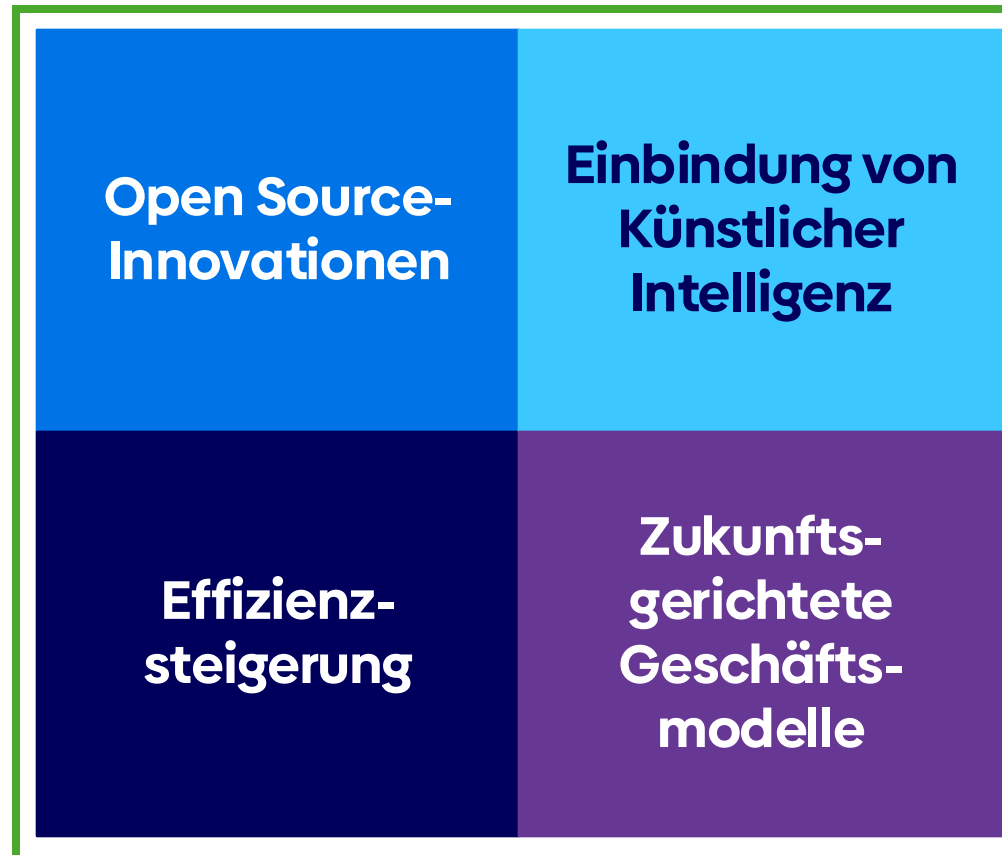
**Effizienzsteigerung**

**Zukunftsgerichtete  
Geschäftsmodelle**



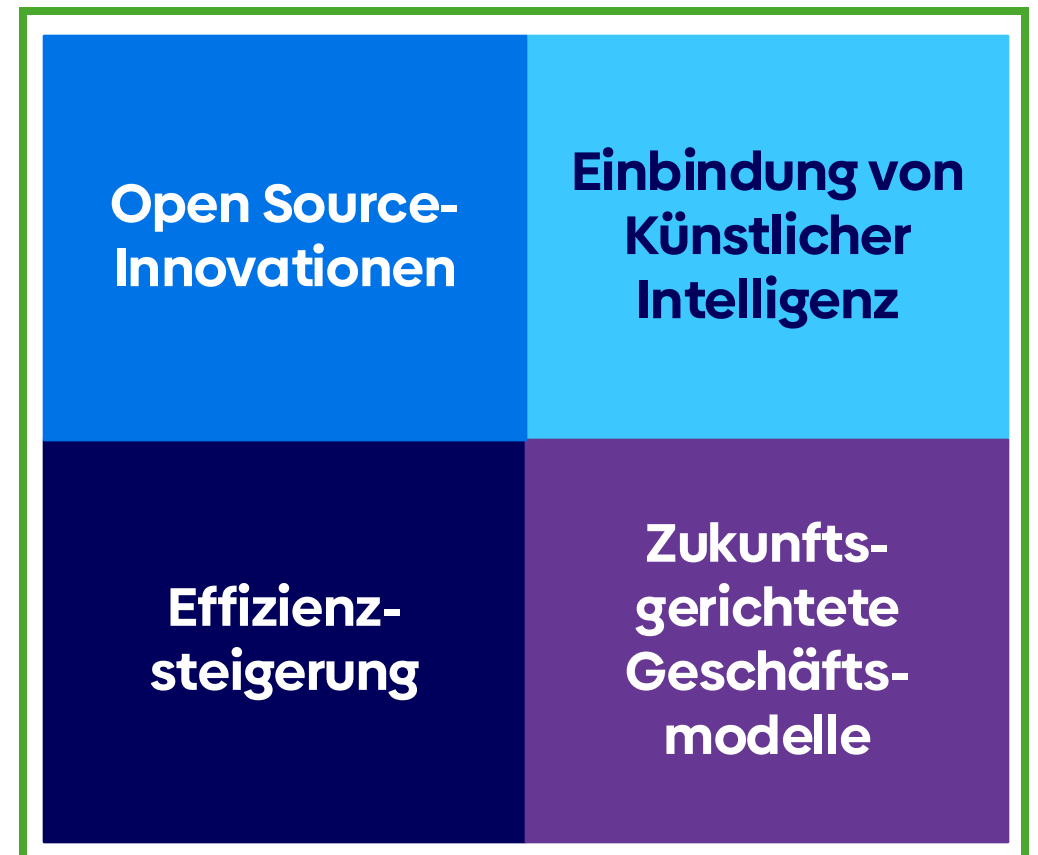
# Die Herausforderung der souveränen Umgebung

Chancen durch neueste Technologien



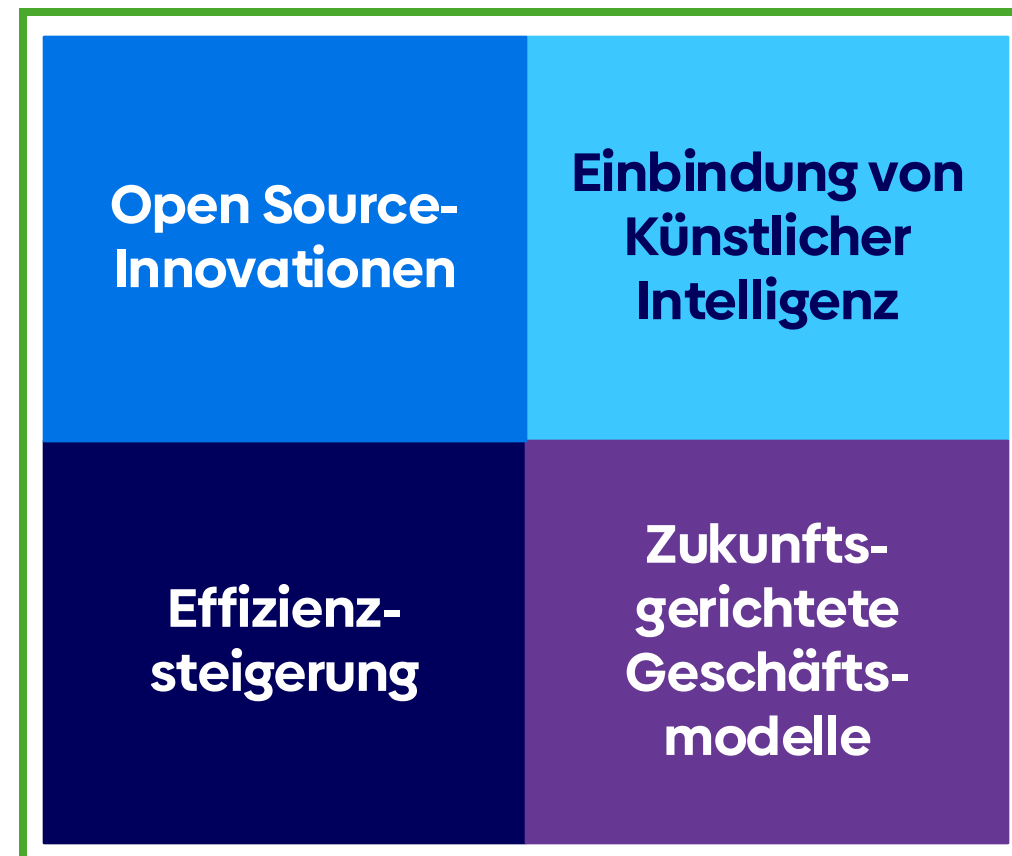
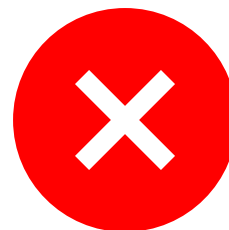
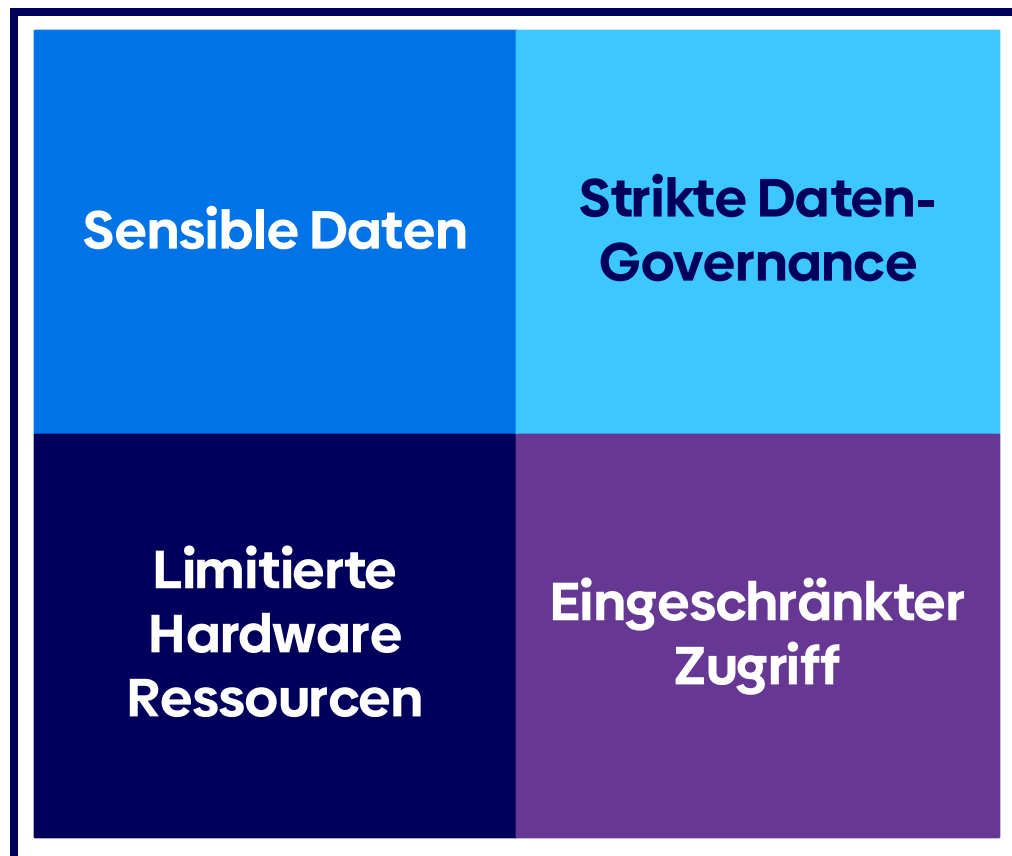
# Die Herausforderung der souveränen Umgebung

Chancen durch neueste Technologien



# Die Herausforderung der souveränen Umgebung

Souveränität vs. effiziente Zukunft



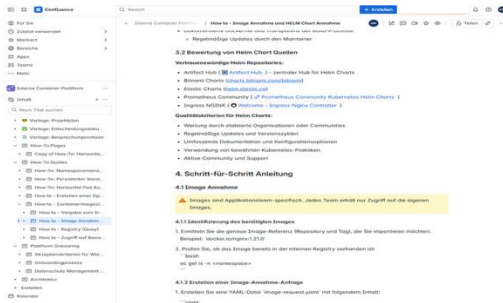
# Die Herausforderung der souveränen Umgebung

## Projektbeispiel aus dem Öffentlichen Sektor – Anforderungen an den Inhalt



### Nutzer der internen Container-Plattform (ICP)

- Nutzer senden **Fragen über die Nutzung** der souveränen Plattform
- Fragt sehr **sensible Daten** an
- Möchte **effizienter werden**



Sendet Fragen

Sendet Antworten



### Spezialisierte ICP Chatbot

- Beantwortet die Fragen von autorisierten Anwendern
- Ist **spezialisiert** auf das spezifische Wissen der gegebenen sensiblen Daten
- **Faktisch richtige** Informationen, mit minimalen Halluzinationen
- **Schnelle Antworten** zur Effizienzsteigerung



# Die Herausforderung der souveränen Umgebung

## Projektbeispiel aus dem Öffentlichen Sektor – Anforderungen an den Inhalt



The screenshot shows the Confluence web interface. On the left is a sidebar with navigation options like 'Für Sie', 'Zuletzt verwendet', 'Markiert', 'Bereiche', 'Apps', 'Teams', and 'Mehr'. Below these are 'Interne Container Plattform' and 'Inhalt'. The main content area displays a document titled 'How to - Image Annahme und HELM Chart Annahme' under the 'Interne Container Plattform' space. The document content includes sections for 'Bewertung von Helm Chart Quellen', 'Vertrauenswürdige Helm Repositories' (listing Artifact Hub, Bitnami Charts, Elastic Charts, Prometheus Community, and Ingress NGINX), 'Qualitätskriterien für Helm Charts', and 'Schritt-für-Schritt Anleitung'. Under '4.1 Image Annahme', there is a yellow warning box stating that images are application-team-specific. Below this, '4.1.1 Identifizierung des benötigten Images' lists steps to identify the image reference and check its availability in the internal registry. '4.1.2 Erstellen einer Image-Annahme-Anfrage' starts with creating a YAML file named 'image-request.yaml'.

# Die Herausforderung der souveränen Umgebung

## Projektbeispiel aus dem Öffentlichen Sektor – Anforderungen an die Sicherheit



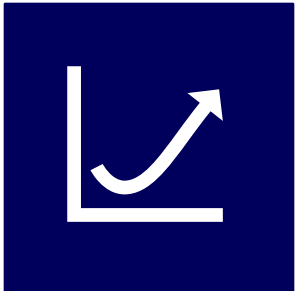
### Datensicherheit

- **Daten** dürfen **sicheren Cluster nicht verlassen**
- Keine Anfragen an öffentliche APIs oder Cloud-Modelle
- Begrenzte Zugriffsrechte



### Transparenz

- **Verhalten** des Chatbots soll möglichst **nachvollziehbar** sein
- Bereitstellung soll über lokal gesteuerte GPU-Cluster erfolgen

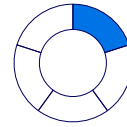


### Skalierbarkeit

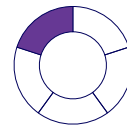
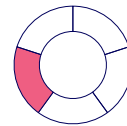
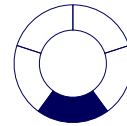
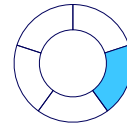
- Modell soll, trotz begrenzter Hardware, skalierbar sein
- **Sichere Ausweitung** auf beliebig viele User & Acceleratoren

# Der Aufbau einer erfolgreichen und sicheren KI-Lösung

## Erfolgsfaktoren



Analysieren, Verstehen und Einbeziehen der Anforderungen an Infrastruktur und Sicherheit



02

# Deep Dive – Language Models



# Deep Dive - Language Models

## Große vs. Kleine Sprachmodelle

### Large Language Models (LLMs)

- Foundation Modelle
- Hohe **generalistische** Performance
- Hoher Ressourcenbedarf zum Trainieren
- **Hoher** Ressourcenbedarf in der Inferenz
- Anfällig für Halluzinationen

### Small Language Models (SLMs)

- **Spezifische** Expertenmodelle
- Performance hoch auf **kleinen Teilgebieten**
- (Mittel-) Hoher Ressourcenbedarf zum Trainieren
- **Niedriger** Ressourcenbedarf in der Inferenz
- **Robuster** gegenüber Halluzinationen

# Deep Dive - Language Models

## Erzeugen von SLM-Expertenmodellen

### RAG

#### Retrieval Augmented Generation

Verbesserung von Gen AI-generierten Texten durch **Abruf relevanter Informationen aus externen Quellen** – für genauere und fundiertere Antworten des Modells.

### InstructLab

#### Large-scale Alignment for chatBots

Nutzung eines **taxonomiegesteuerten** Prozesses zur **synthetischen Datengenerierung** und eines **mehrstufigen Tuning-Frameworks**, um die Modellleistung zu verbessern.

### Finetuning

#### Finetuning

**Anpassung eines vortrainierten Modells** an spezifische Aufgaben oder Daten, um Leistung und Genauigkeit für spezialisierte Anwendungen zu verbessern – ohne vollständiges Neutraining.

# Deep Dive - Language Models

## Erzeugen von SLM-Expertenmodellen

### RAG

#### Retrieval Augmented Generation

Verbesserung von Gen AI-generierten Texten durch **Abruf relevanter Informationen aus externen Quellen** – für genauere und fundiertere Antworten des Modells.

### InstructLab

#### Large-scale Alignment for chatBots

Nutzung eines **taxonomiegesteuerten** Prozesses zur **synthetischen Datengenerierung** und eines **mehrstufigen Tuning-Frameworks**, um die Modellleistung zu verbessern.

### Finetuning

#### Finetuning

**Anpassung eines vortrainierten Modells** an spezifische Aufgaben oder Daten, um Leistung und Genauigkeit für spezialisierte Anwendungen zu verbessern – ohne vollständiges Neutraining.

# Deep Dive - Language Models

## Der InstructLab-Workflow



# Deep Dive - Language Models

## Vorteile von SLMs und InstructLab



### Spezialisierung und Verständnis

- Verbesserte Performance auf den gewünschten Themengebieten
- Das **Modell lernt und versteht**, statt nur auswendig zu lernen



### Transparenz

- Kontrolliertes und transparentes Lernen durch den Einsatz einer Taxonomie
- Beeinflussen der erlernten Inhalte
- **Verringern von Halluzinationen**



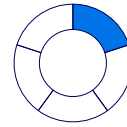
### Effizienz

- Das finale Modell ist klein und leichtgewichtig
- Ermöglicht **schnellere Antworten**
- **Reduziert Ressourcenbedarf** im produktiven Einsatz

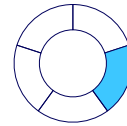


# Der Aufbau einer erfolgreichen und sicheren KI-Lösung

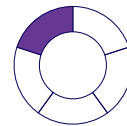
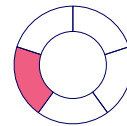
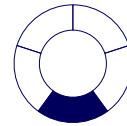
## Erfolgsfaktoren



Analysieren, Verstehen und Einbeziehen der Anforderungen an Infrastruktur und Sicherheit



Nutzen spezialisierter Expertenmodelle



03

# Deep Dive – OpenShift AI Plattform

# Deep Dive – OpenShift AI Plattform

## Kernbestandteile von OpenShift AI



### Data Science Projects:

MLOps auf OpenShift

Kubernetes Integration

Abstraktion der Infrastruktur



### Workbenches:

u.a. Jupyter Notebooks

Kollaborative  
Entwicklung

Cluster GPU  
Unterstützung

Experimentieren



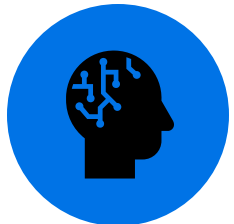
### Data Science Pipelines:

Codeausführung  
(sequentiell, parallel)

Zentrale  
Orchestrierung

Skalierbarkeit

Direkte OpenShift  
Pipeline Integration



### Model Serving:

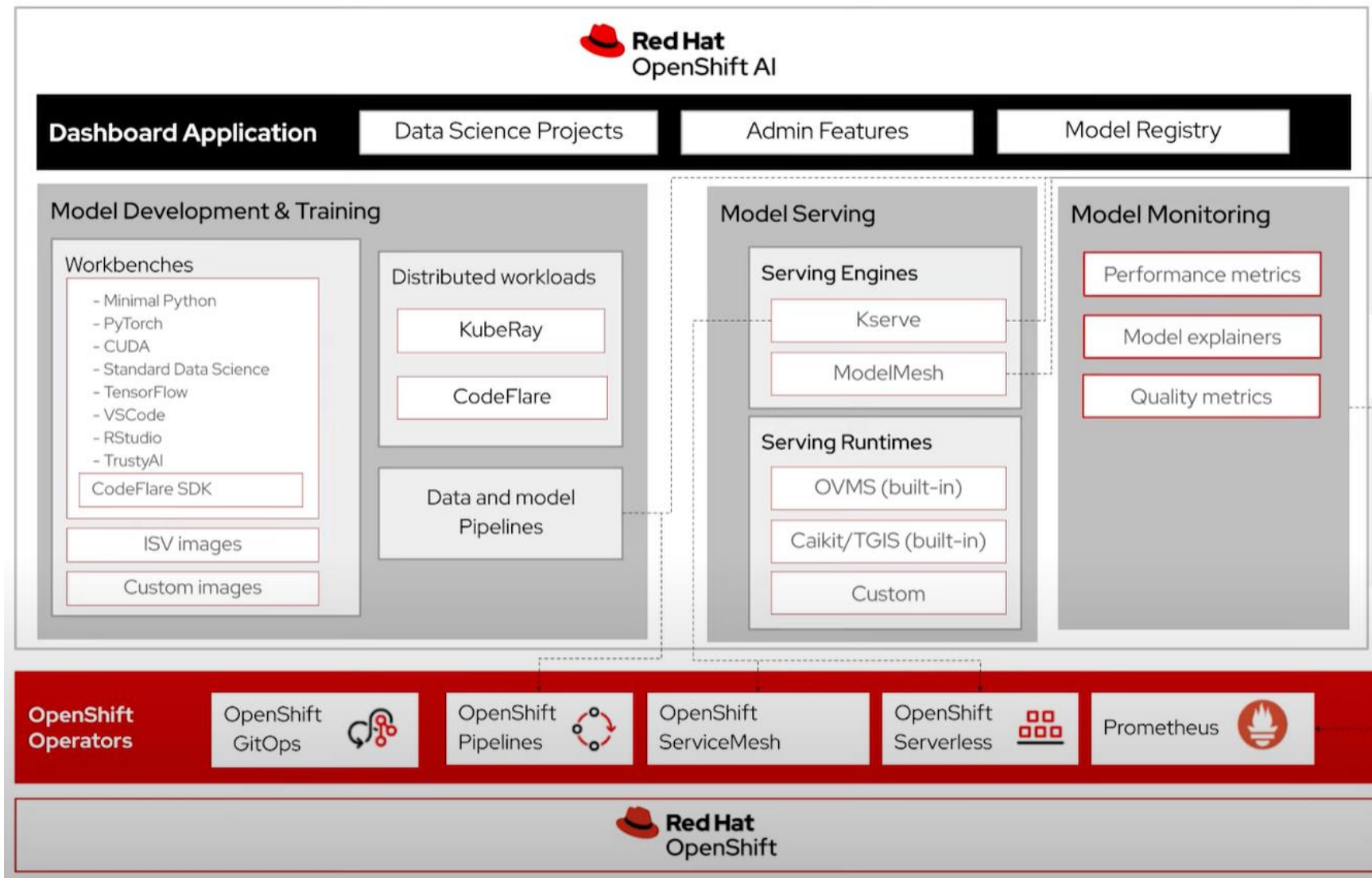
Schnelle Bereitstellung

Autoscaling

Monitoring

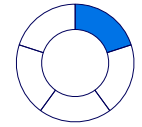
# Deep Dive – OpenShift AI Plattform

## Technischer Hintergrund von OpenShift AI

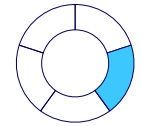


# Der Aufbau einer erfolgreichen und sicheren KI-Lösung

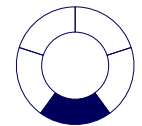
## Erfolgsfaktoren



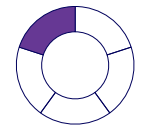
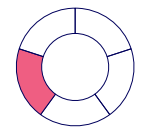
Analysieren, Verstehen und Einbeziehen der Anforderungen an Infrastruktur und Sicherheit



Nutzen spezialisierter Expertenmodelle



Integration einer ganzheitlichen AI Management Plattform



04

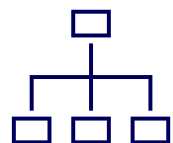
# Überblick zur Lösungs- architektur



# Lösungsarchitektur

## Workflow-Überblick

Verarbeitung großer  
Eingangsdaten (**ICP-  
Confluence**) mit Docling  
zur Umwandlung in MD-  
Dateien



Finetuning wird **mehrfach  
wiederholt**, wobei die Taxonomie  
bei jeder Iteration **verbessert** wird

InstructLab nutzt  
Taxonomie, um  
**ähnliche Frage-  
Antwort-Beispiele  
zu erzeugen**, mit  
denen das LLM  
lernt



**InstructLab**

Bereitstellung des  
angepassten  
Modells auf einem  
**vLLM-Modellserver**  
innerhalb von  
OpenShift AI



**Streamlit**

**Datenquelle**

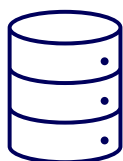
**Taxonomie**

**InstructLab**

**Angepasstes  
Modell**

**Deployment**

**Nutzung**



Taxonomie  
**strukturiert** die  
Informationen der  
Eingangsdaten in  
**Wissenszweige**  
und ein **Frage-  
Antwort-Format**



**InstructLab**

Nach dem  
Training ist das  
Modell **auf** die  
bereitgestellten  
**spezifischen  
Informationen  
angepasst**



**Red Hat  
OpenShift AI**

Verwendung des  
Modells in einer  
Streamlit-  
**Anwendung**, die  
Anfragen über  
einen **Endpoint** an  
das Modell sendet

# Lösungsarchitektur

## Zusammenarbeit mit Red Hat

### Modell-Training

#### Atos

- **Übermittelt Taxonomie** an spezialisiertes **InstructLab-Team** bei Red Hat

#### Red Hat

- nutzt die **Rechenleistung von NVIDIA GPUs**, um das InstructLab-Modell zu trainieren



Kontinuierlicher  
Feedback-  
Loop zwischen  
Atos & Red Hat



### Modell-Deployment und Validierung

#### Atos

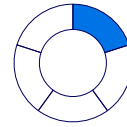
- **Modell-Deployment** auf OpenShift AI
- **Evaluierung der Modell-Performance**
- Taxonomie-Anpassung / Optimierung
- Taxonomie an Red Hat zurückgeben für neue Trainingsiteration

→ Dieser **Prozess wird wiederholt**, bis das Modell die gewünschte Leistung erreicht

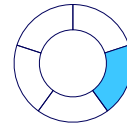


# Der Aufbau einer erfolgreichen und sicheren KI-Lösung

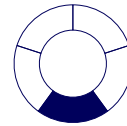
## Erfolgsfaktoren



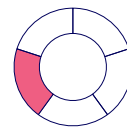
Analysieren, Verstehen und Einbeziehen der Anforderungen an Infrastruktur und Sicherheit



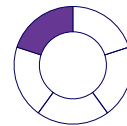
Nutzen spezialisierter Expertenmodelle



Integration einer ganzheitlichen AI Management Plattform



Entwickeln eines maßgeschneiderten Implementierungs-Konzepts

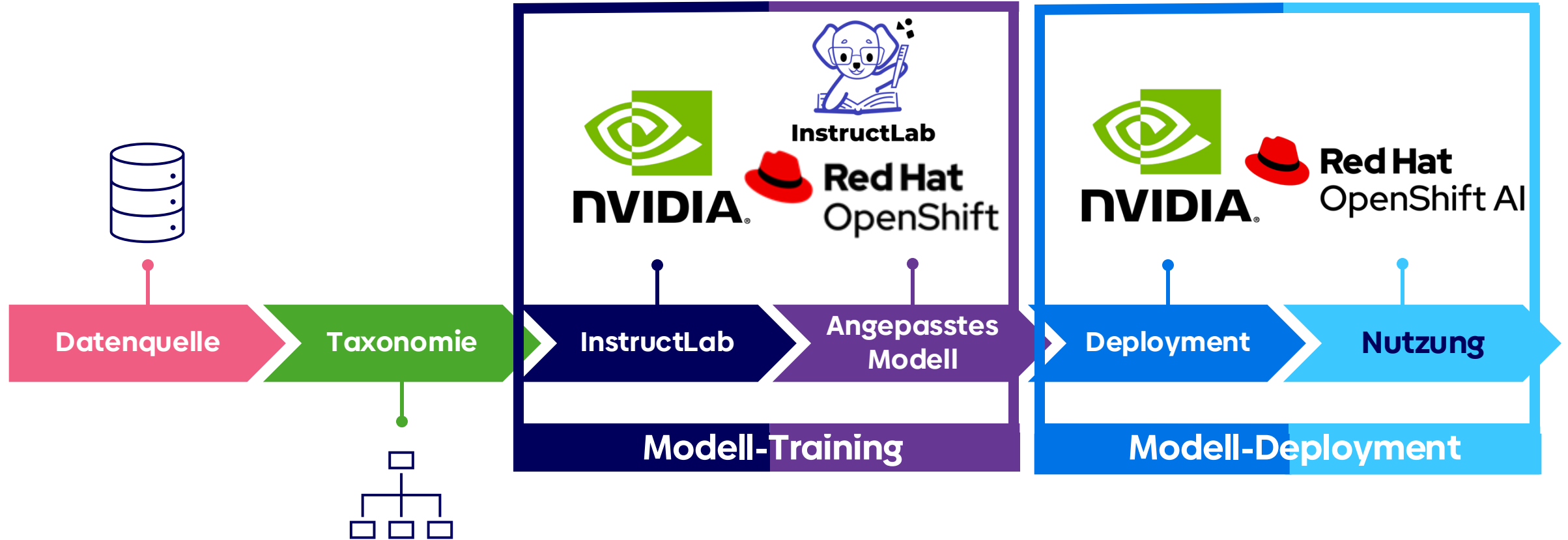


05

# Modell-Training

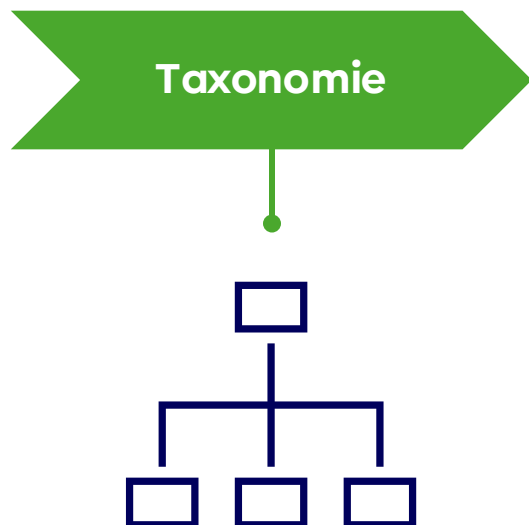
# Modell-Training

## Genutzte Infrastruktur



# Modell-Training

## Taxonomie-Einblick

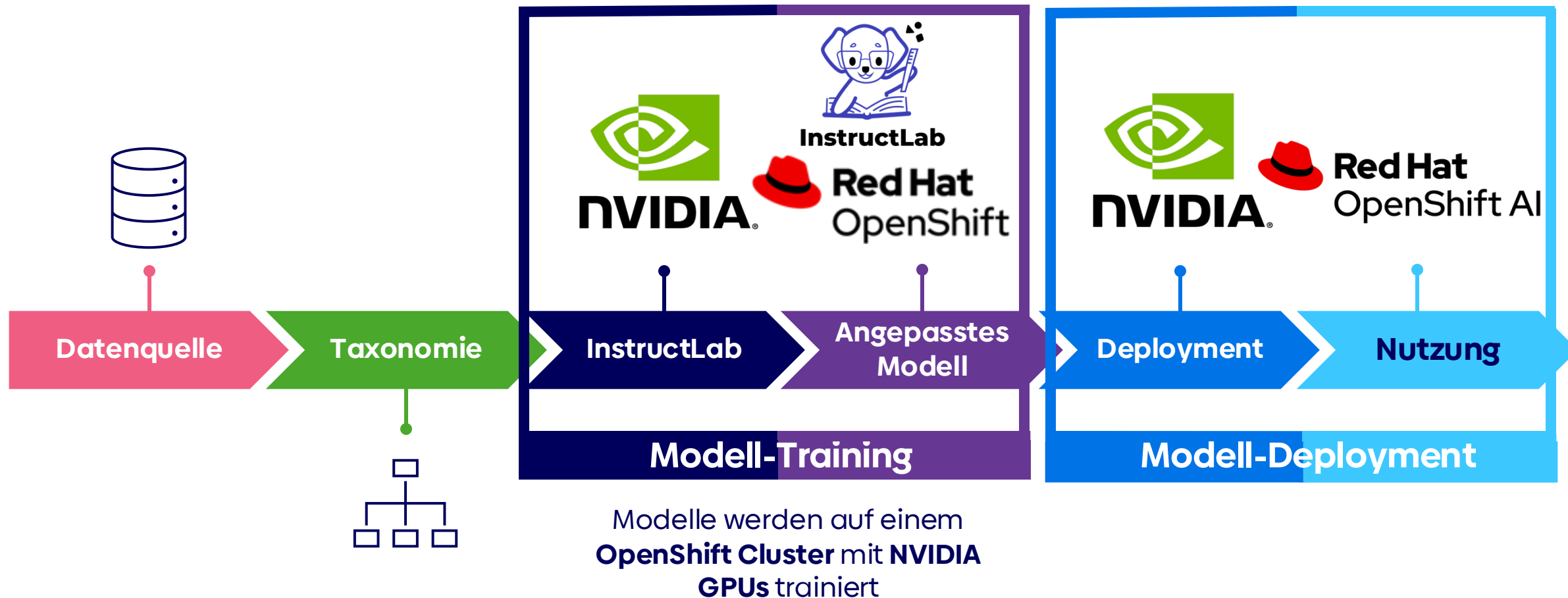


```
1  ---
2  version: 3
3  domain: Container Management
4  created_by: jm02k02
5  document_outline: |
6      Informationen zum Handhaben von Container Images und Helm Charts auf der
7      unternehmensinternen Containerplattform.
8  seed_examples:
9      - context: |
10          ## 2. Voraussetzungen
11          - Gültiges Benutzerkonto für die OpenShift Container Plattform
12          - Zugehörigkeit zu einem Applikationsteam mit entsprechenden
13            Berechtigungen
14          - Grundlegende Kenntnisse über Container und Helm Charts
15          - Zugang zum internen Registry-Service
16          - Installierte OpenShift CLI (`oc`)
17          - Optional: Installierte Helm CLI für lokale Tests
18
19  questions_and_answers:
20      - question: |
21          Welche Sicherheitskriterien müssen nach den Vorgaben unserer
22          Plattform für die Bewertung von Container Images verwendet werden?
23        answer: |
24          Die folgenden Sicherheitskriterien sind von unserem Plattformteam zur
25          Bewertung vorgegeben.
```



# Modell-Training

## Genutzte Infrastruktur

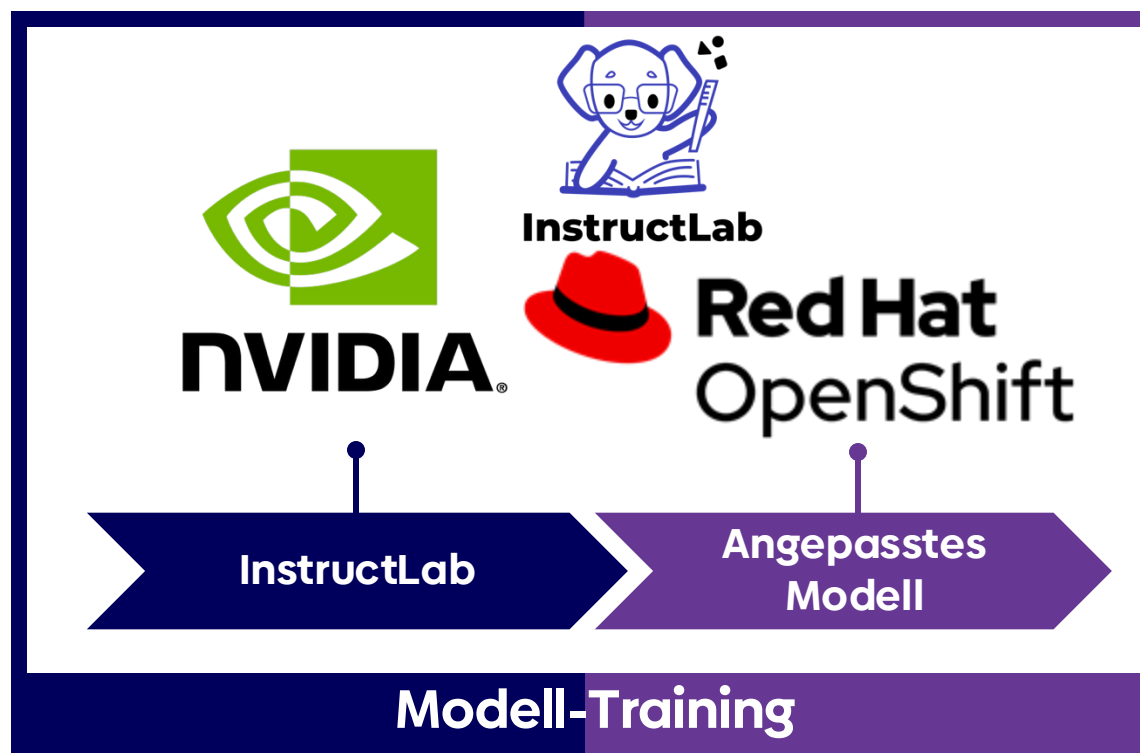


# Modell-Training

## Trainings-Infrastruktur der ersten Iterationen

Aufbau einer  
InstructLab custom  
pipeline für deutsche  
Sprache

Hardware für SDG:  
4 NVIDIA L40S



Teacher-Modell:  
Llama 3.3 70B  
Validator-Modell:  
---

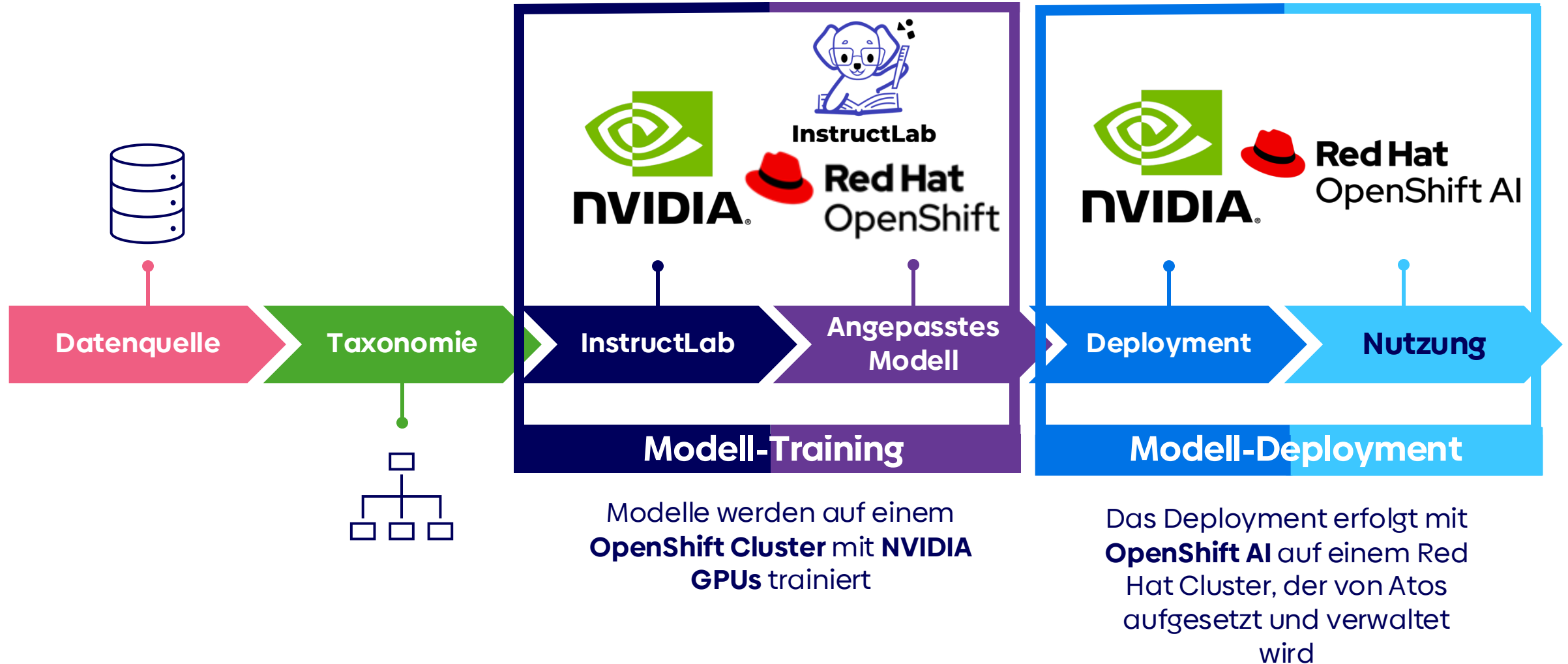
Student-Modell:  
granite-3.1-8b-starter-  
v1

06

# Modell- Deployment

# Modell-Deployment

## Genutzte Infrastruktur

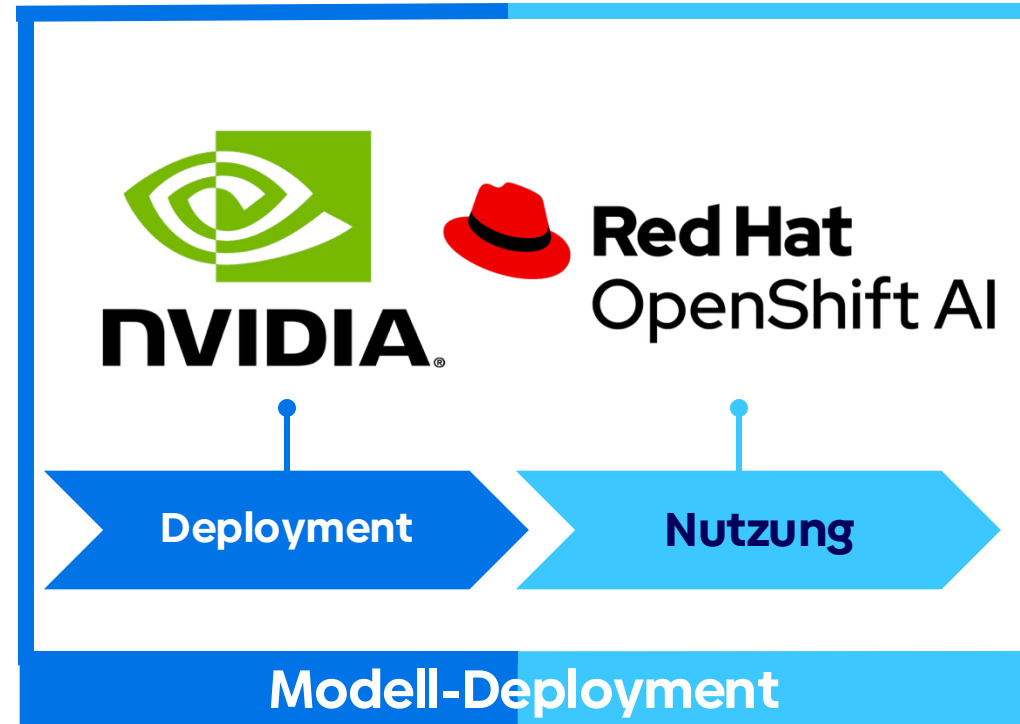


# Modell-Deployment

## Details zum Deployment auf OpenShift AI

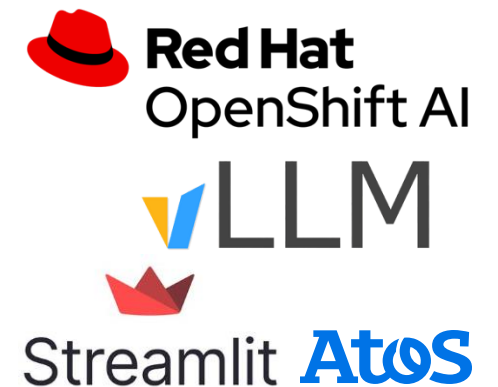
vLLM-Interaktion über  
S3-bucket oder über  
model car-Container

Eine GPU-worker node:  
Standard\_NC12s\_v3  
(2 Tesla-V100-PCIE-  
16GB)



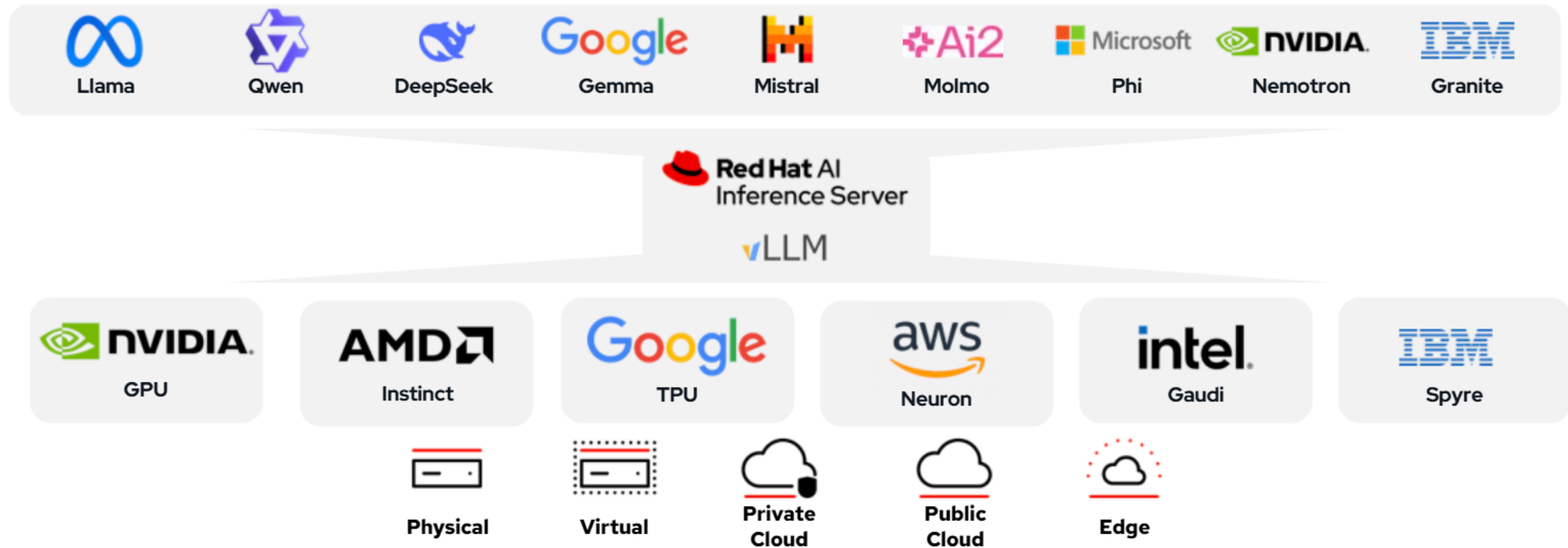
OpenShift AI (v 2.22.0)  
workbenches zum  
Definieren der  
Endpoints

Streamlit spricht vLLM-  
model server an  
Version:  
vllm:rhoai-2.20-cuda



# Modell-Deployment

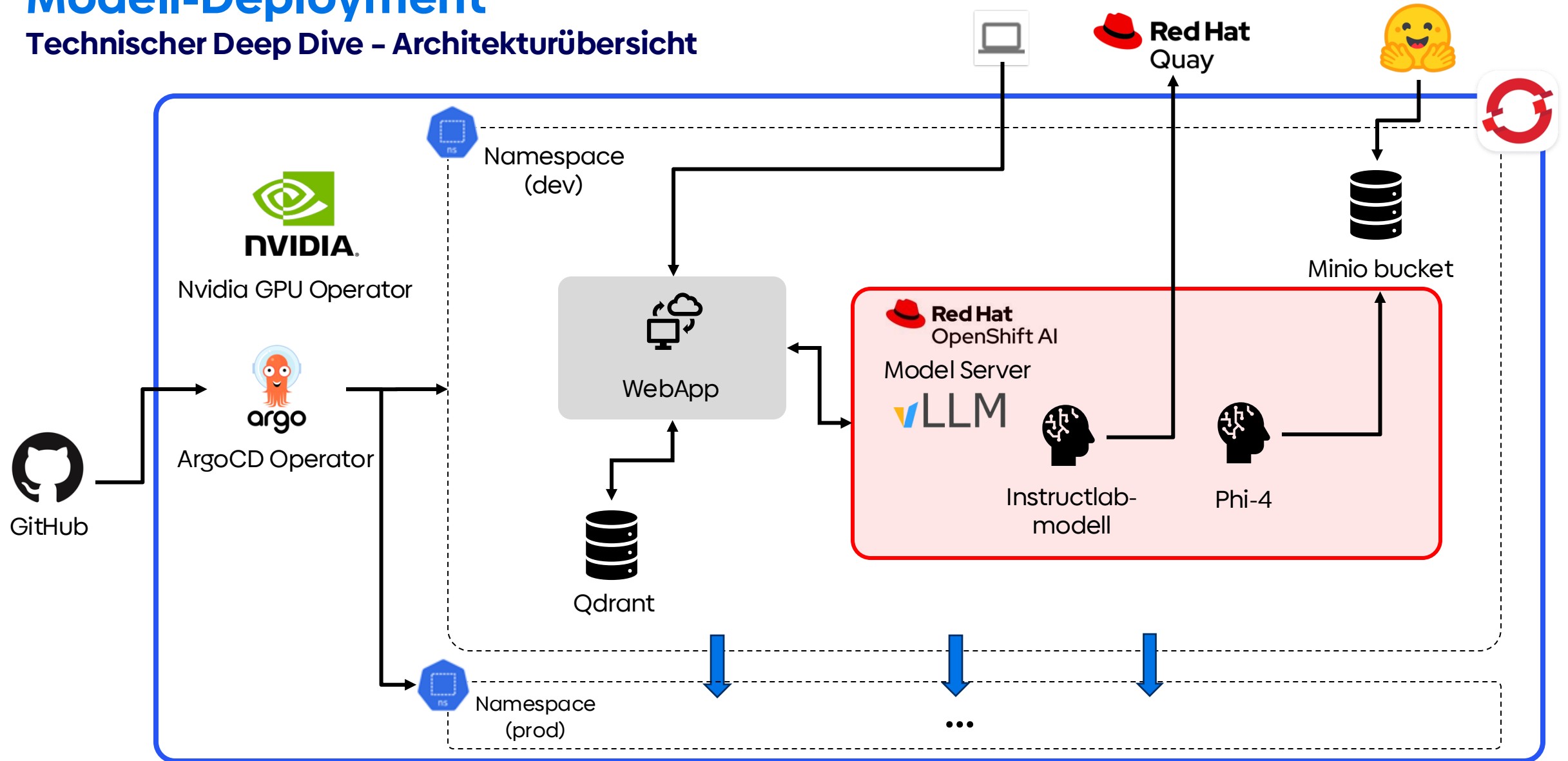
## vLLM: Der Kern des LLM-Deployments





# Modell-Deployment

## Technischer Deep Dive – Architekturübersicht

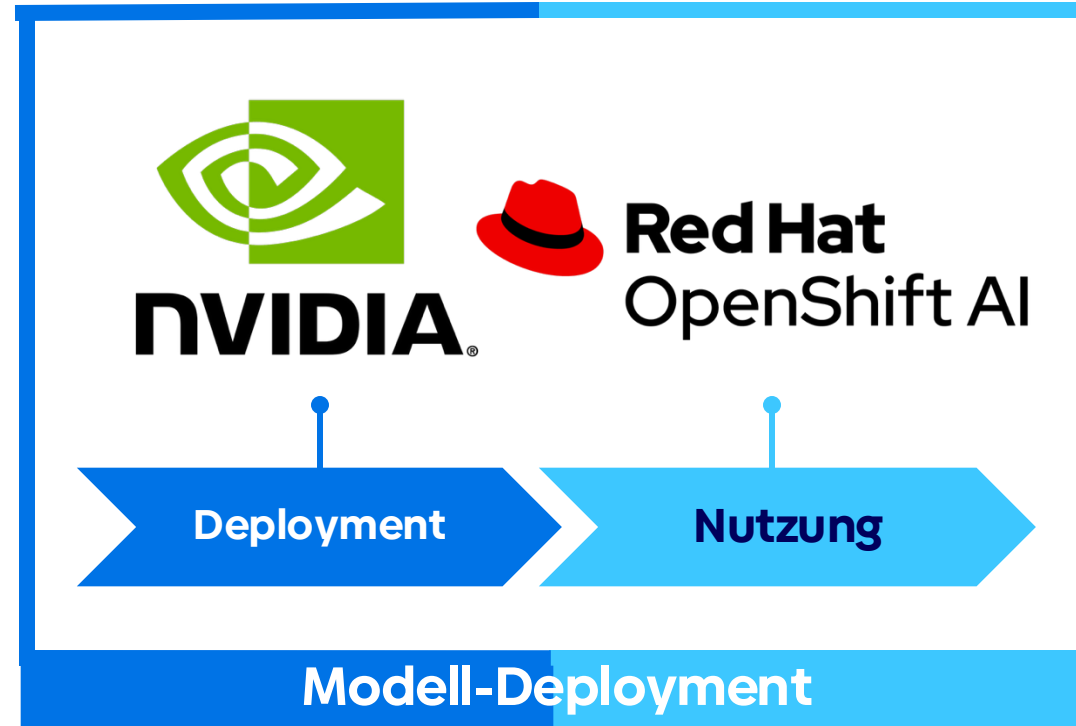


# Modell-Deployment

## Details zum Deployment auf OpenShift AI

vLLM-Interaktion über  
S3-bucket oder über  
model car-Container

Eine GPU-worker node:  
Standard\_NC12s\_v3  
(2 Tesla-V100-PCIE-  
16GB)

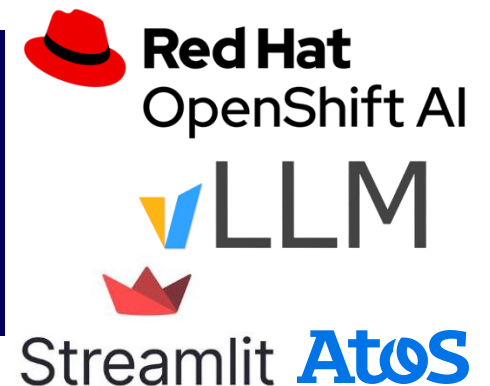


OpenShift AI (v 2.22.0)  
workbenches zum  
Definieren der  
Endpoints

Streamlit spricht vLLM-  
model server an  
Version:  
vllm:rhoai-2.20-cuda

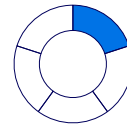


- Zusammenfassung:
- **Keine** manuelle Arbeit auf **einzelnen Containern**
  - Deployment ist **hochverfügbar** und einfach zu konfigurieren
  - Nutzung ist **voll managed, supported und automatisiert**
  - **Jedes AI-Modell** trainierbar und nutzbar

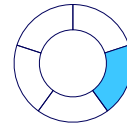


# Der Aufbau einer erfolgreichen und sicheren KI-Lösung

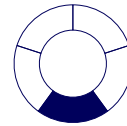
## Erfolgsfaktoren



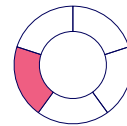
Analyse und Einbeziehung der Infrastruktur-Voraussetzungen



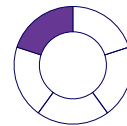
Nutzen spezialisierter Expertenmodelle



Integration einer ganzheitlichen AI Management Plattform



Entwickeln eines maßgeschneiderten Implementierungs-Konzepts



Plattform-agnostisches Deployment

**07**

# **Modell- evaluierung**

# Modellevaluierung

## Zusammensetzung der Metrik

**Festes Set an Validierungsfragen  
(~ 12 pro Wissenszweig)**

**4 Modelle werden evaluiert:**

- Phi4 Base-Modell
- Getunted InstructLab-Modell
- Phi4 + RAG
- Getunted InstructLab + RAG



**Der Modell-  
Score**

**Jeder Antwort jedes Modells wird ein  
Score von 1 (sehr schlecht) bis 5  
(Ausgezeichnet) vergeben**

**Alle Teammitglieder bewerten  
isoliert**  
→ repräsentativ und  
unvoreingenommen  
→ Mittelwert aller Wertungen ist  
finaler Modell-Score

# Modellevaluierung

## Beispielfrage 1

Welche Labels sind für die Metrik `icp_quota_utilization_ratio` definiert ?

**Phi4-Base: 1 / 5**

Um Informationen über spezifische Metriken wie ``icp_quota_utilization_ratio`` zu erhalten, einschließlich der damit verbundenen Labels, müsste man in der Regel auf die Dokumentation oder die technischen Spezifikationen der Plattform oder des Tools zugreifen, das diese Metrik definiert.

**InstructLab: 1 / 5**

Die Labels für die Metrik `icp_quota_utilization_ratio` sind nicht explizit im Dokument definiert. Die Labels könnten jedoch wie folgt aussehen:

- ``icp_metric``: `"icp_quota_utilization_ratio"`
- ``icp_type``: `"gauge"`
- ``icp_description``: `"Ratio der Quota-Auslastung"`
- ``icp_unit``: `"ratio"`

**Phi4-RAG: 5 / 5**

Die Metrik ``icp_quota_utilization_ratio`` hat die folgenden Labels definiert:

- ``namespace`` und
- ``resource_type``.

**InstructLab-RAG:  
5 / 5**

Die Metrik ``icp_quota_utilization_ratio`` ist mit den Labels ``namespace`` und ``resource_type`` definiert.

# Modellevaluierung

## Beispielfrage 2

Was ist der Link für die Interne Plattform Dokumentation?

**Phi4-Base: 1 / 5**

Um dir beim Zugriff auf die interne Plattformdokumentation zu helfen, benötige ich spezifische Informationen über die Plattform oder das Unternehmen, dem du angehörst. Normalerweise findest du solche Links in internen Unternehmensressourcen, wie zum Beispiel einem Intranet, einer Mitarbeiter-App oder durch direkte Kommunikation von deinem Team oder IT-Support.

**InstructLab: 5 / 5**

Der Link zur internen Plattform Dokumentation lautet: ``https://confluence.internal.example.com/container-platform``

**Phi4-RAG: 1 / 5**

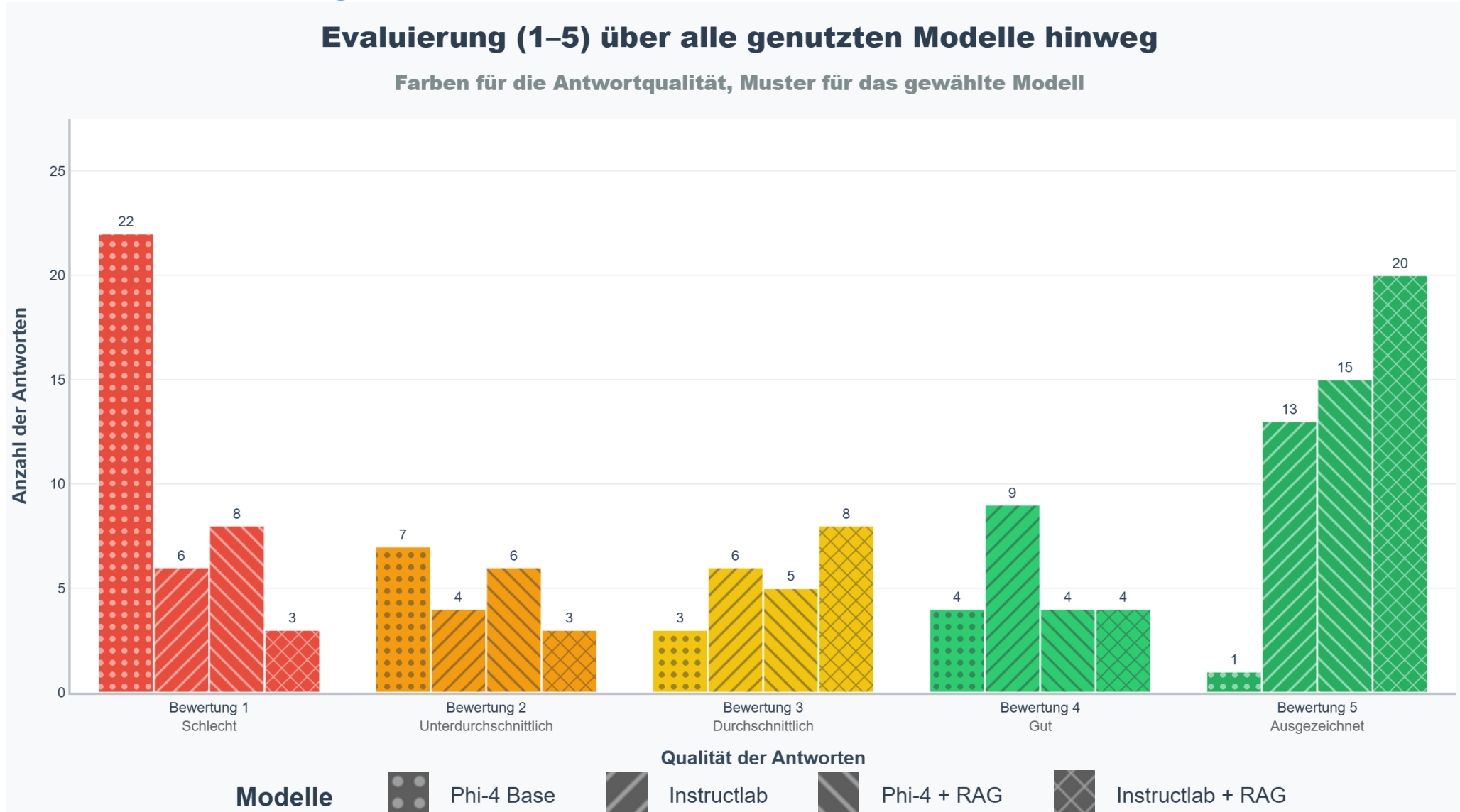
Basierend auf dem bereitgestellten Kontext enthält der Text keinen spezifischen Link zur Internen Plattform Dokumentation. Um den Link zu finden, empfehle ich, sich direkt an das Plattformteam zu wenden, das unter `plattform-support@icp-company.de` erreichbar ist. Sie können dort um Zugang zur vollständigen Dokumentation oder weitere Informationen bitten.

**InstructLab-RAG: 5 / 5**

Der Link für die Interne Plattform Dokumentation ist nicht explizit im bereitgestellten Dokument angegeben, jedoch wird er im Kontext als "Die Interne Container-Plattform (ICP)" erwähnt, die eine Dokumentation enthält. Der genaue Link lautet: `https://confluence.internal.example.com/container-platform/`.



# Modellevaluierung






# Modellevaluierung

## Durchschnittliche Ergebnisse im Überblick

	Phi4 Base	InstructLab Base	Phi4 RAG	InstructLab RAG
Container Management	2	4,3	3,4	3,9
Plattform Architektur	1,5	2,6	2,5	3,5
Plattform Onboarding	1,8	3,3	3,1	2,8
Gesamt	1,7	3,3	3	3,4

# Modellevaluierung

## Durchschnittliche Ergebnisse im Überblick

	Phi4 Base	InstructLab Base	Phi4 RAG	InstructLab RAG
Gesamt	/	Beste Nutzbarkeit	Geringste time-to-value	Beste Performance
				
		<ul style="list-style-type: none"><li>- Wissen steht sofort zur Verfügung</li><li>- Nur ein model server benötigt</li></ul>	<ul style="list-style-type: none"><li>- Kein Training benötigt</li><li>- Modellgröße unverändert</li></ul>	<ul style="list-style-type: none"><li>- Kombiniert Stärken von InstructLab und RAG</li><li>- Sehr transparent</li></ul>

08

# Fazit und Ausblick

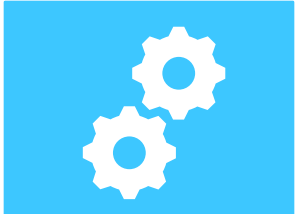
# Fazit und Ausblick

## Integration von KI in bestehende souveräne Cluster



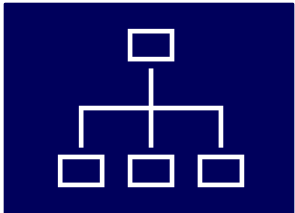
### Sicheres Training und Deployment von Künstlicher Intelligenz

- Beibehalten der eigenen **Datensouveränität**
- **Volle Kontrolle** über Training und Anwendung der KI



### OpenShift AI als ganzheitliche KI-Plattform

- Anknüpfung **an bestehende containerisierte souveräne Applikationen**
- Bringt Skalierbarkeit, Monitoring und Self-Healing in die Welt von KI
- Erlaubt Integration **neuester Technologien in eine sichere Infrastruktur**



### InstructLab als Tool zur Erstellung von SLMs

- Kleine **transparente Expertenmodelle**, statt großer Generalistenmodelle
- Angepasst auf interne Informationen des Unternehmens
- Geringe Größe erlaubt **lokales Deployment auf limitierter Hardware**

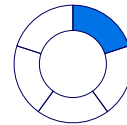


### Effizienzgewinne

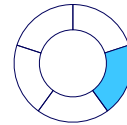
- Integration von innovativen KI-Lösungen in bestehende sichere Umgebungen
- **Schneller Zugang zu sehr genauen Ergebnissen**
- Liefert **zielgerichteten Mehrwert** für die Beschäftigten im Arbeitsalltag

# Der Aufbau einer erfolgreichen und sicheren KI-Lösung

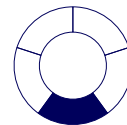
## Erfolgsfaktoren



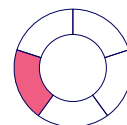
Analysieren, Verstehen und Einbeziehen der Anforderungen an Infrastruktur und Sicherheit



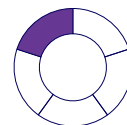
Nutzen spezialisierter Expertenmodelle



Integration einer ganzheitlichen AI Management Plattform



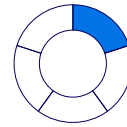
Entwickeln eines maßgeschneiderten Implementierungs-Konzepts



Plattform-agnostisches Deployment

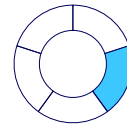
# Der Aufbau einer erfolgreichen und sicheren KI-Lösung

## Erfolgsfaktoren



Analysieren, Verstehen und Einbeziehen der Anforderungen an Infrastruktur und Sicherheit → **Atos**

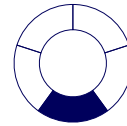
**Atos**



Nutzen spezialisierter Expertenmodelle → **InstructLab**

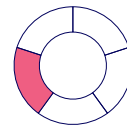


**InstructLab**



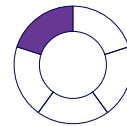
Integration einer ganzheitlichen AI Management Plattform → **OpenShift AI**

 **Red Hat**  
OpenShift AI



Entwickeln eines maßgeschneiderten Implementierungs-Konzepts → **Atos**

**Atos**



Plattform-agnostisches Deployment → **OpenShift AI**

 **Red Hat**  
OpenShift AI



# OpenShift AI, InstructLab und Atos

Die Brücke zwischen der sicheren Umgebung und Künstlicher Intelligenz







# Fragen & Diskussion

# Danke !

Bei Rückfragen, gerne melden ! ☺

[Marius.Kiskemper@atos.net](mailto:Marius.Kiskemper@atos.net)  
[Justus.mrosk@atos.net](mailto:Justus.mrosk@atos.net)



Atos is a registered trademark of Atos SE. © 2024 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.

# Atos