



Connect Large Language Modelle und der Models-as-a-Service Ansatz

Selbst ausprobieren, sicher managen und Kosten senken

Anne Faulhaber

Senior Technical Account Manager
Red Hat

Manuel Hahn

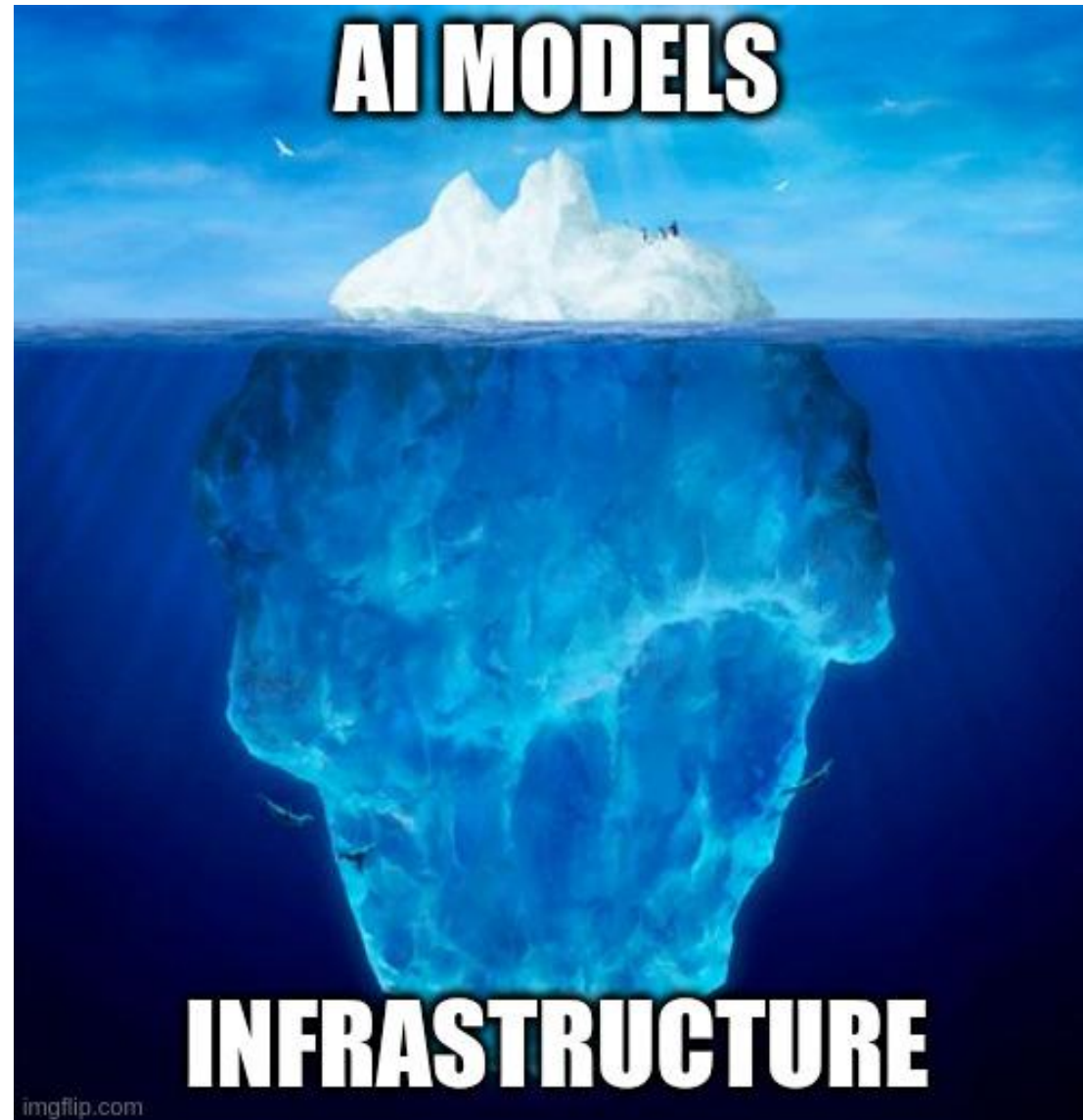
Associate Principal Specialist Solution Architect
Red Hat



Dieser Workshop zeigt, wie "Models as a Service" (MaaS) auf OpenShift, AI Hardware besser nutzt und Kosten senkt.

- Inhalte:
 - a. Einführung in die OpenShift AI-Plattform.
 - b. Sichere Bereitstellung von Generative AI-Modellen via API Management Tool.
 - c. Konsumierung über ein API Gateway (z.B. für einen AI Coding Assistenten).
 - d. Nutzung von Agentic AI/MCP in Chat-Anwendungen.
 - e. Erfassung von Nutzungsstatistiken.
- Ziel: Sie beherrschen die effiziente und sichere MaaS-Bereitstellung von LLMs.





Common Concerns of Gen AI Projects

The Case for Model Hosting



Data privacy

Some data can't leave the organization.

Host models in secure environment.



Cost

Real-world applications incur significant costs.

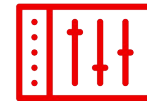
Control costs by using own hardware. Maximize utilization and efficiency!



Adaptability

Can't customize models to boost use-case expertise.

Fine-tune and host models.

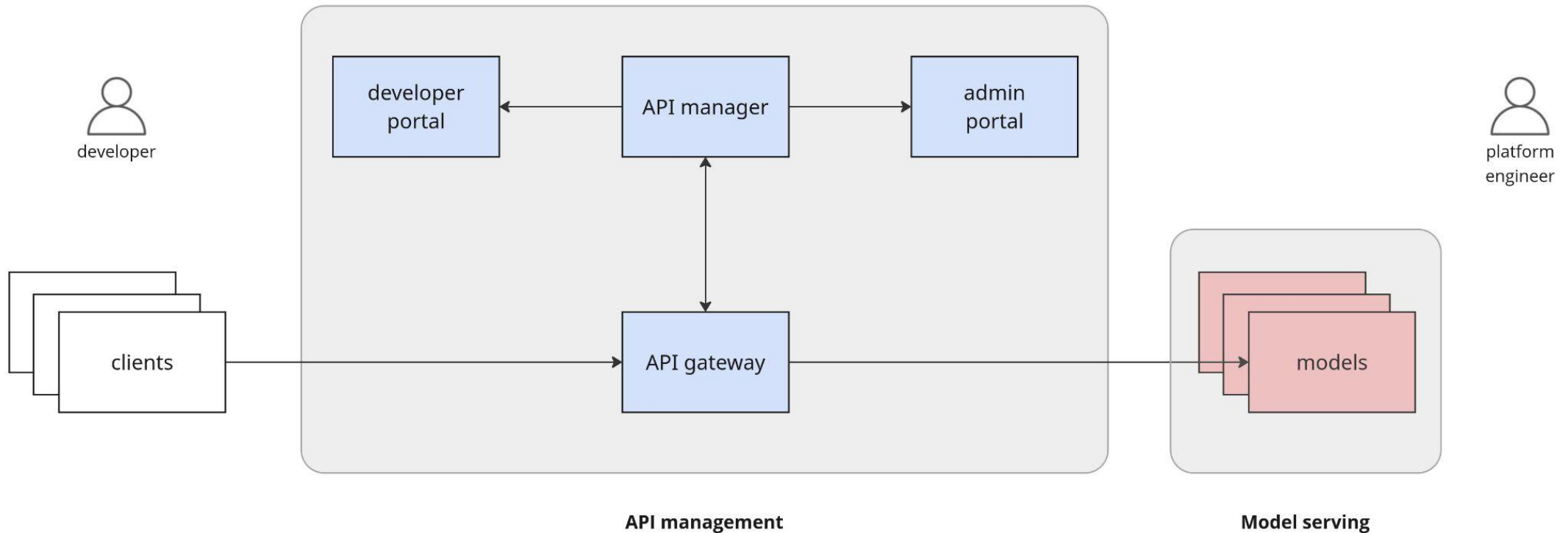


Control

Lack of transparency regarding training data and model versions.

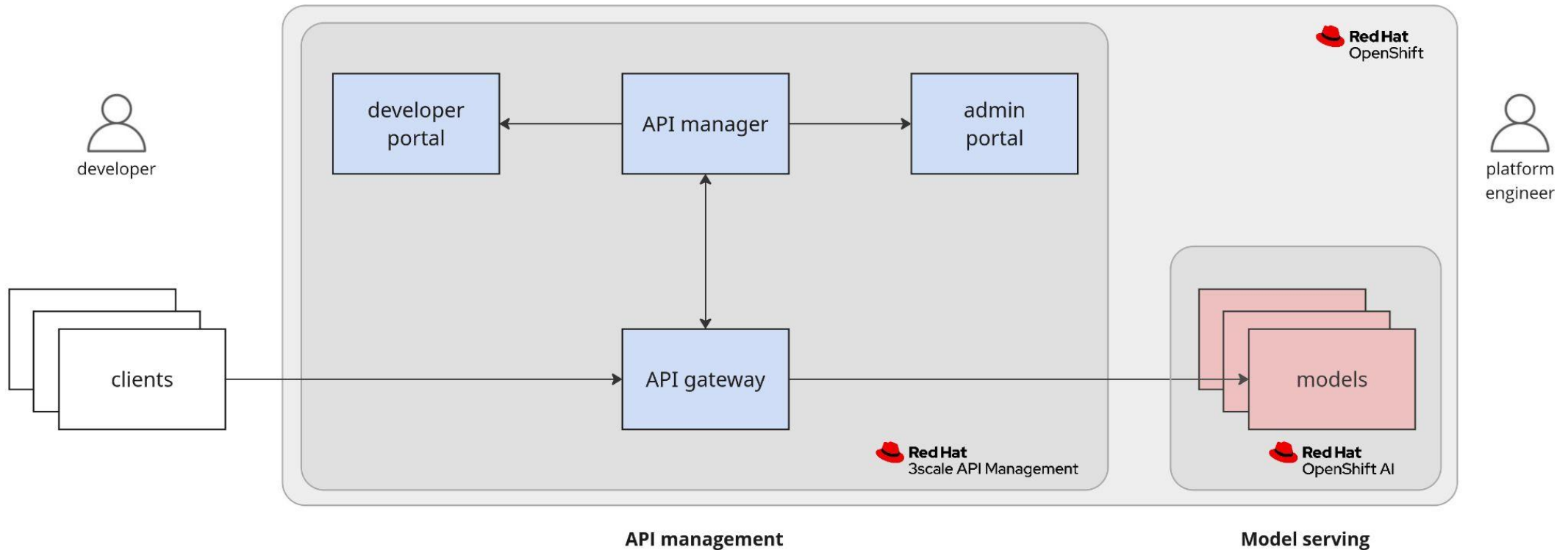
Host open source model and control deployment lifecycle.

Models as a Service – Architecture

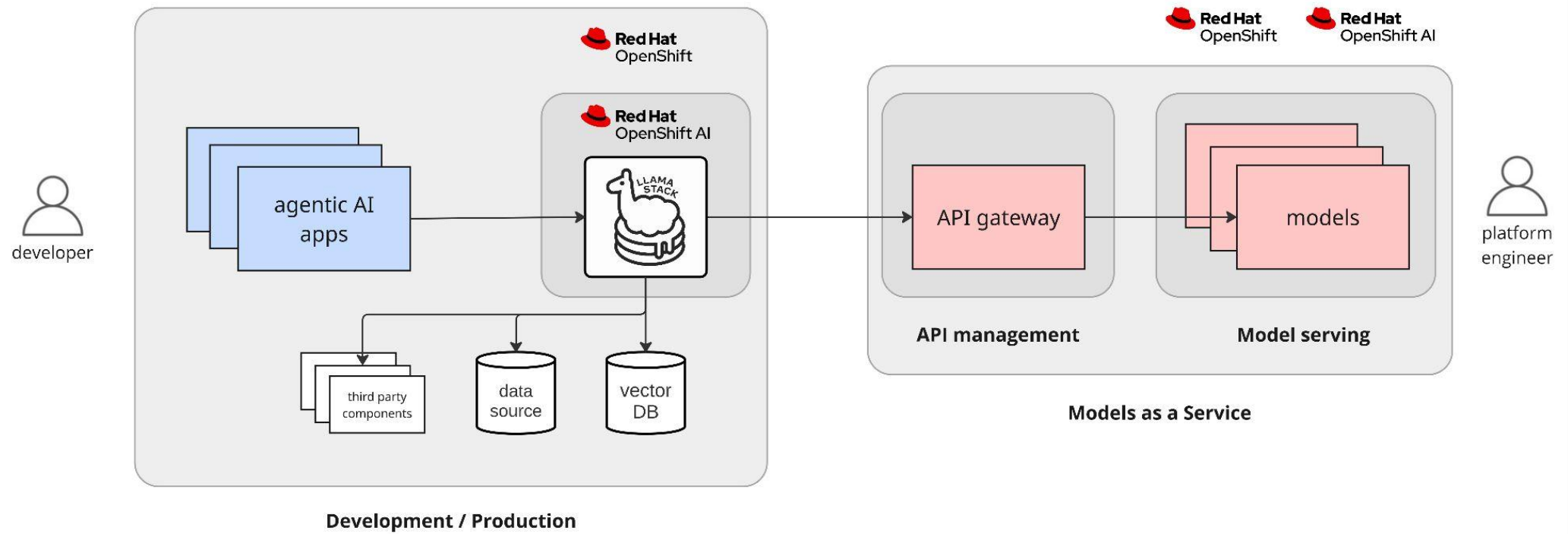


Models as a Service – Architecture

A Sample Implementation

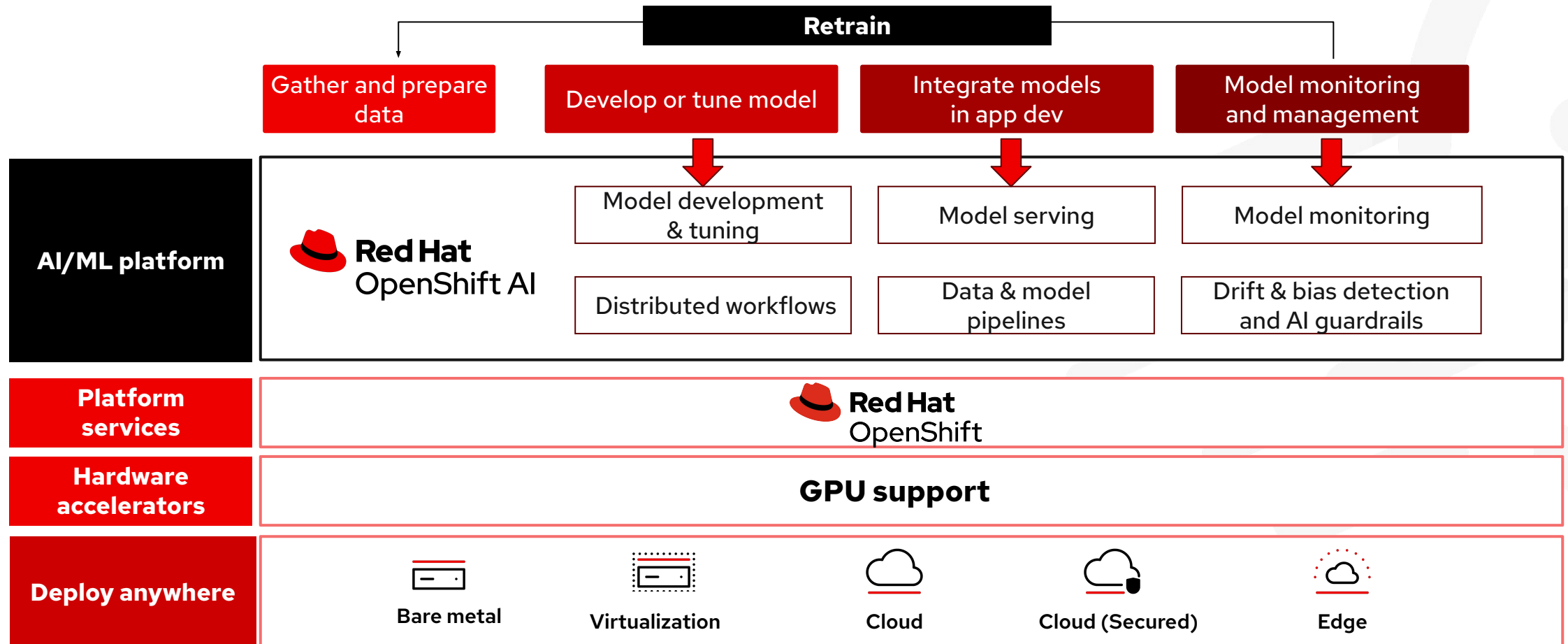


Models as a Service & Agentic AI



Red Hat OpenShift AI

One platform for generative and predictive AI



Lab Dashboard

All you need!

Open the following website to access your lab environment:

red.ht/llm-as-a-service

Password: aimodels



Permanent lab guide:

<https://rhpds.github.io/private-summit-llmaas-showroom>





Jetzt Session bewerten!

Einfach QR-Code scannen,
Session aus der Liste wählen
und bewerten. **Vielen Dank!**

red.ht/rhsc-darmstadt-feedback



Connect

Thank you



linkedin.com/company/red-hat



facebook.com/redhatinc



youtube.com/user/RedHatVideos



twitter.com/RedHat

