Overview of Red Hat OpenShift AI

MLOps platform for artificial intelligence/ machine learning (AI/ML) use cases

Red Hat

Operationalizing AI/ML is not trivial

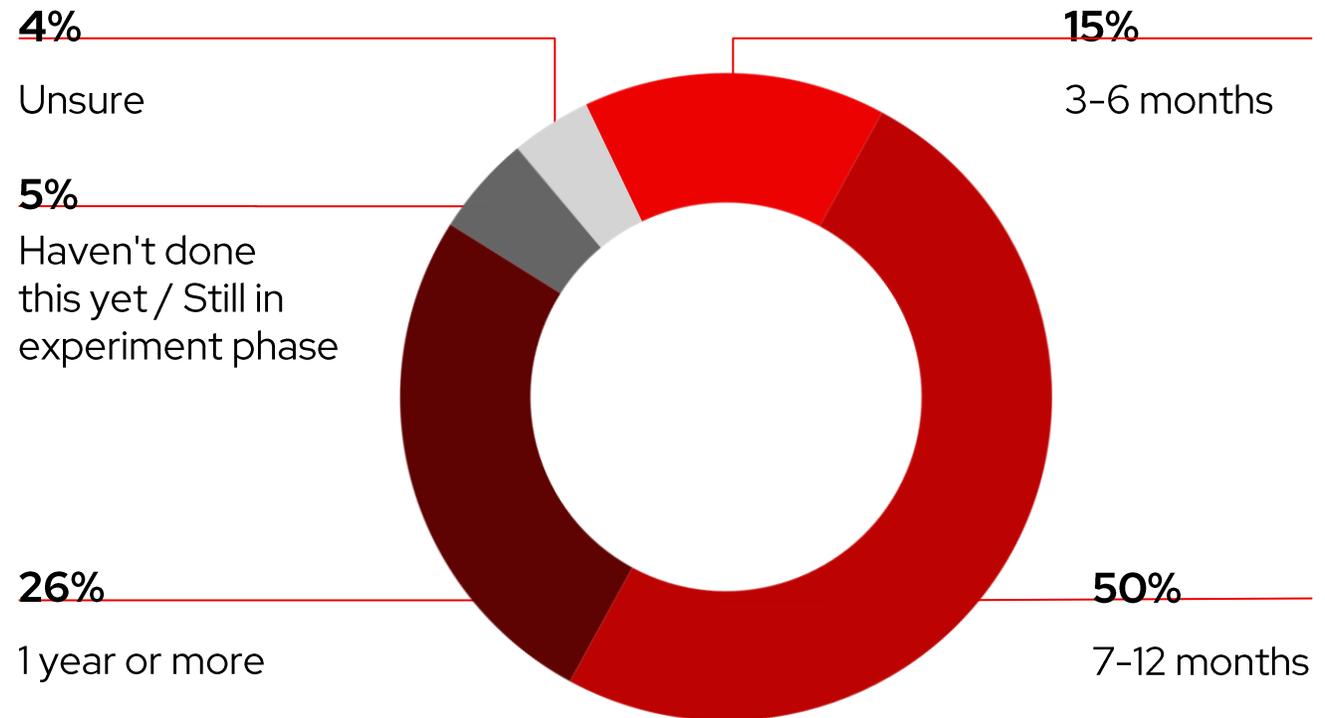## Every member of your team plays a critical role in a complex process

	Set goals	Gather and prepare data	Develop model	Integrate models in app dev	Model monitoring and management
Business leadership	━━━━━━━━				
Data engineer		━━━━━━━━			
Data scientist			━━━━━━━━		━━━━━━━━
ML engineer			━━━━━━━━━━━━━━━━━━━━━━━━━━━		
App developer				━━━━━━━━━━━━━━	
IT operations		━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━			

Red Hat

Operationalizing AI is still a challenging process

What is the average AI/ML timeline from idea to operationalizing the model?

Half of respondents (50%) say their average AI/ML timeline from idea to operationalizing the model is 7–12 months.

4%
Unsure

5%
Haven't done this yet / Still in experiment phase

26%
1 year or more

15%
3–6 months

50%
7–12 months

3

Red Hat's AI portfolio strategy

Trust

Choice

Consistency

AI models

RHEL AI

Base Model | Alignment Tuning | Methodology & Tools | Platform Optimization & Acceleration

AI platform

OpenShift AI

Development | Serving | Monitoring & Lifecycle | MLOps | Resource Management

AI enabled portfolio

Lightspeed portfolio

Usability & Adoption | Guidance | Virtual Assistant | Code Generation

AI workload support

Optimize AI workloads

Deployment & Run | Compliance | Certification | Models | Open Source Ecosystem

Open Hybrid Cloud Platforms

Red Hat Enterprise Linux | Red Hat OpenShift | Red Hat Ansible Platform
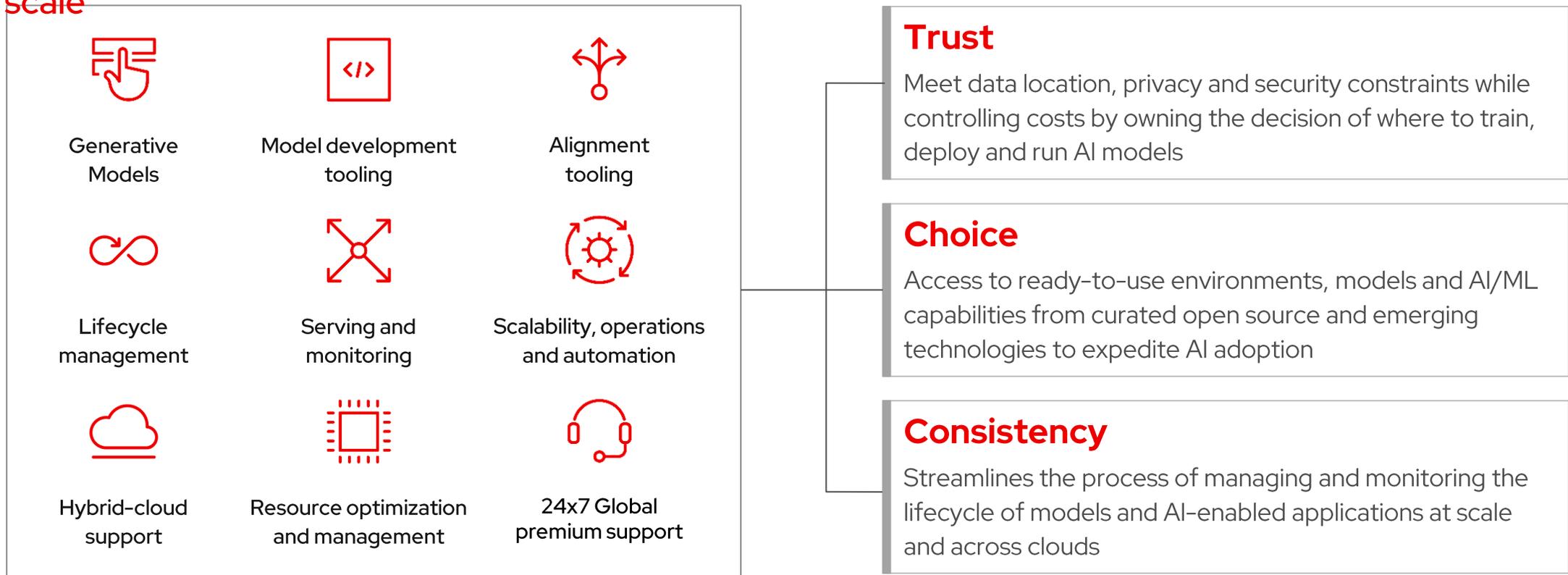
Acceleration | Performance | Scale | Automation | Observability | Security | Developer Productivity | App Connectivity | Secure Supply Chain

Partner Ecosystem

Hardware | Accelerators | Delivery

Red Hat

Red Hat AI platforms

Red Hat offers generative AI and MLOps capabilities for building flexible, trusted AI solutions at scale

Generative Models

Model development tooling

Alignment tooling

Lifecycle management

Serving and monitoring

Scalability, operations and automation

Hybrid-cloud support

Resource optimization and management

24x7 Global premium support

Trust

Meet data location, privacy and security constraints while controlling costs by owning the decision of where to train, deploy and run AI models

Choice

Access to ready-to-use environments, models and AI/ML capabilities from curated open source and emerging technologies to expedite AI adoption

Consistency

Streamlines the process of managing and monitoring the lifecycle of models and AI-enabled applications at scale and across clouds

Red Hat

Red Hat AI platforms

Red Hat Enterprise Linux AI

Foundation model platform for developing, testing, and running Granite family LLMs

▸ Provides a simplified approach to get started with generative AI that includes open source models

▸ Makes AI accessible to developers and domain experts with little data science expertise

▸ Provides the ability to do training & inference on individual production server deployments

Red Hat OpenShift AI

Integrated MLOps platform for model lifecycle management at scale anywhere

▸ Provides support for both generative and predictive AI models with a BYOM approach

▸ Includes distributed compute, collaborative workflows, model serving and monitoring

▸ Offers enterprise MLOps capabilities and the ability to scale across hybrid–clouds

▸ Includes Red Hat Enterprise Linux AI, including the Granite family models

Red Hat strategy around generative AI and foundation models

▸ Provide Granite LLM models and tooling for model customization as part of RHEL AI and OpenShift AI

▸ Developing the infrastructure stack for single server and distributed workloads, scheduling for building, fine-tuning and serving foundation models

▸ Enable out-of-the-box "bring your own model" use cases to OpenShift AI to incorporate open source and partner ecosystem model integration

▸ Red Hat is infusing Generative AI capabilities into more of its portfolio.

- OpenShift Lightspeed preview and RHEL Lightspeed vision announced at Red Hat Summit

Red Hat's AI/ML engineering is 100% open source

Integrated AI platform

Create and deliver gen AI and predictive models at scale across hybrid cloud environments.

Available as
- Fully managed cloud service
- Traditional software product on-site or in the cloud!

Model development
Bring your own models or customize Granite models to your use case with your data. Supports integration of multiple AI/ML libraries, frameworks, and runtimes.

Model serving and monitoring
Deploy models across any OpenShift footprint and centrally monitor their performance.

Lifecycle management
Expand DevOps practices to MLOps to manage the entire AI/ML lifecycle.

Resource optimization and management
Scale to meet workload demands of gen AI and predictive models. Share resources, projects, and models across environments.
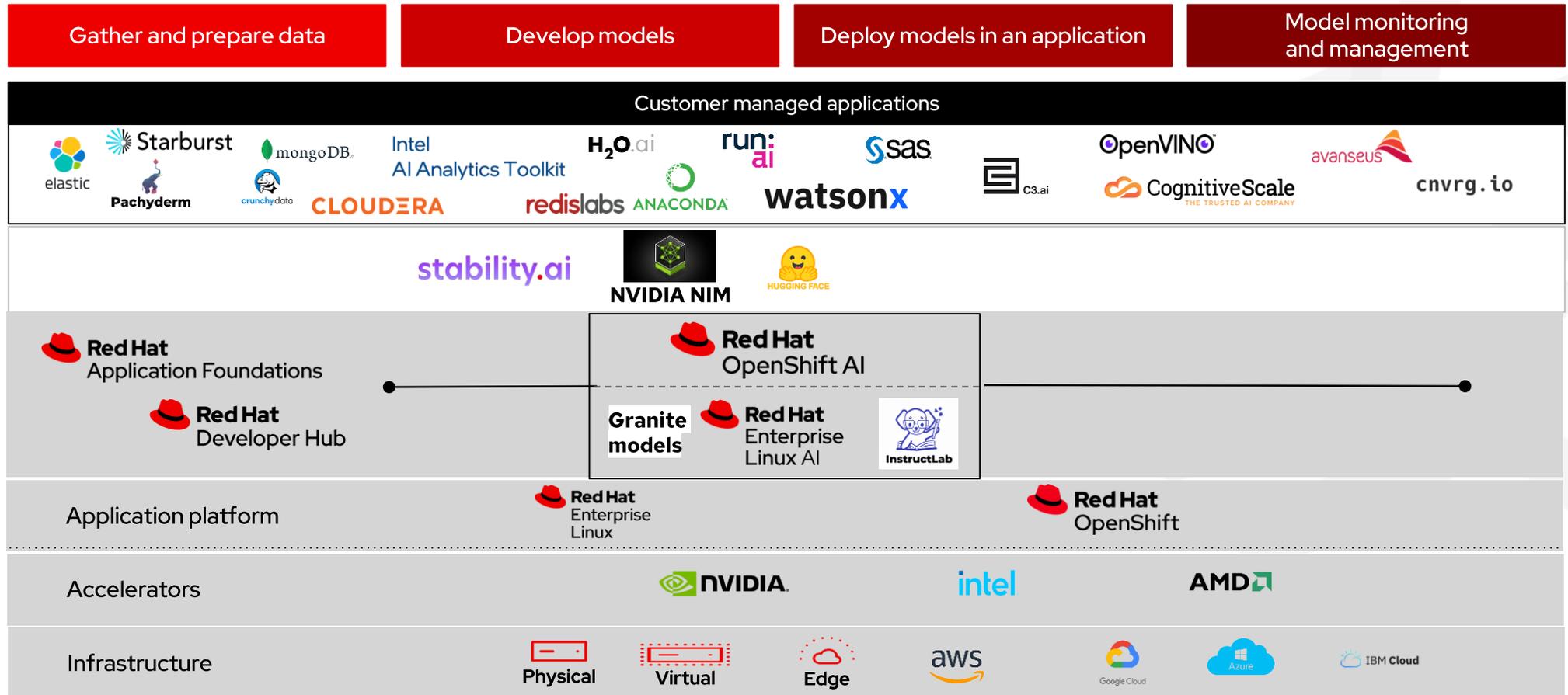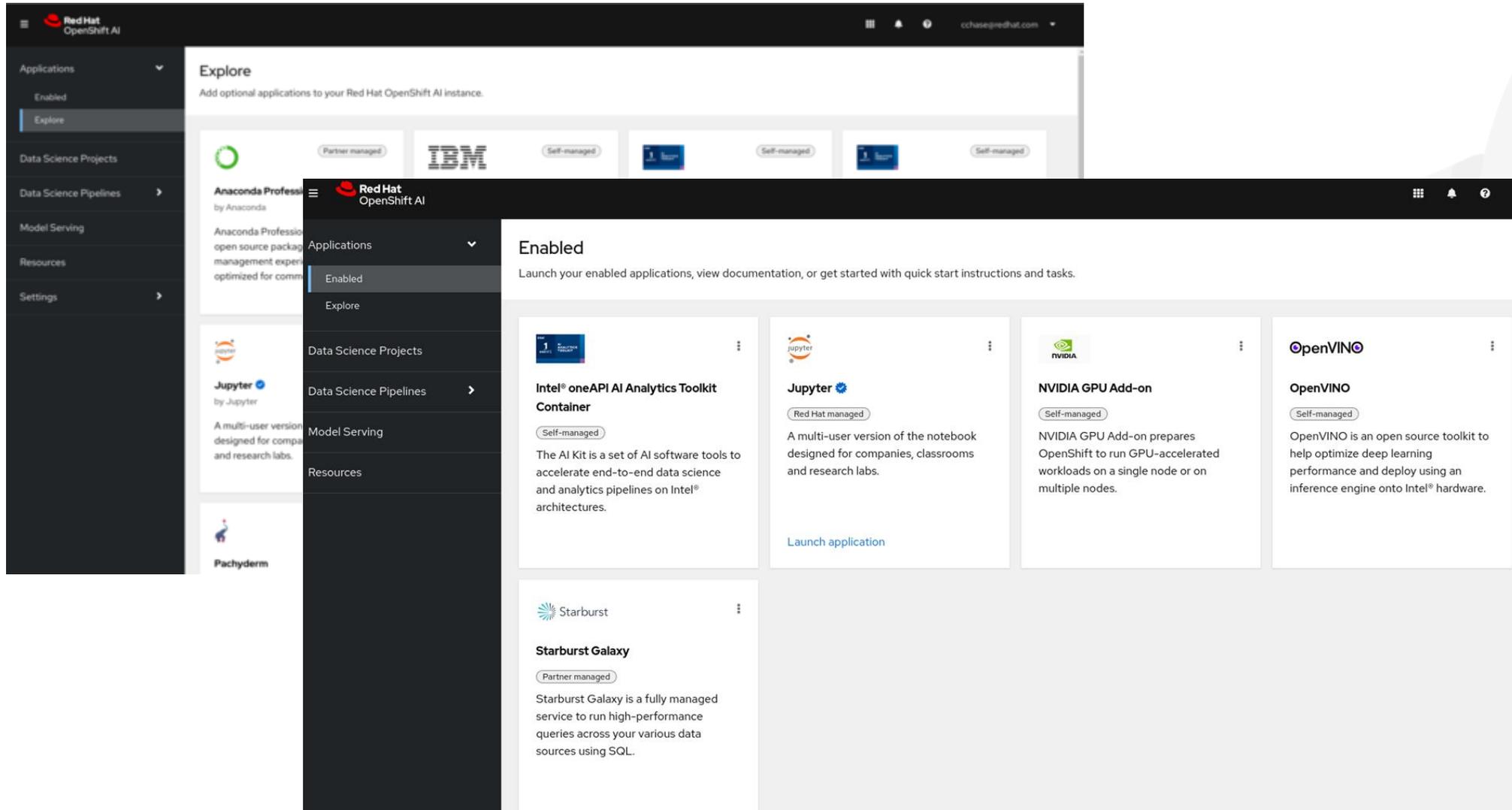
OpenShift AI components

Retrain

| Gather and prepare data | Develop models | Deploy models in an application | Model monitoring and management |

watsonx

ISV software and cloud services

Starburst
Pachyderm elastic

intel AI TOOLS
ANACONDA

intel OpenVINO

Red Hat software and cloud services

Red Hat Application Foundations

Data streaming | Integration

Red Hat Enterprise Linux AI

jupyter PyTorch
Distributed workloads

Model serving | Data science pipelines | Model performance monitoring

Red Hat on-premise and cloud platform

Red Hat Enterprise Linux   Red Hat OpenShift

Accelerators: NVIDIA intel AMD

On-premise, cloud or edge infrastructure

Red Hat

Detailed look integrating our partner ecosystem

Gather and prepare data	Develop models	Deploy models in an application	Model monitoring and management

Customer managed applications

ISV software and services

elastic · Starburst · Pachyderm · mongoDB · crunchy data · Intel AI Analytics Toolkit · CLOUDERA · redislabs · ANACONDA · H2O.ai · run:ai · watsonx · SAS · C3.ai · OpenVINO · CognitiveScale THE TRUSTED AI COMPANY · avanseus · cnvrg.io

3rd party models and repositories

stability.ai · NVIDIA NIM · HUGGING FACE

Red Hat software and cloud services

Red Hat Application Foundations

Red Hat Developer Hub

Red Hat OpenShift AI

Granite models · Red Hat Enterprise Linux AI · InstructLab

Red Hat cloud platform

Application platform — Red Hat Enterprise Linux · Red Hat OpenShift

Accelerators — NVIDIA · intel · AMD

Infrastructure — Physical · Virtual · Edge · aws · Google Cloud · Azure · IBM Cloud

Red Hat

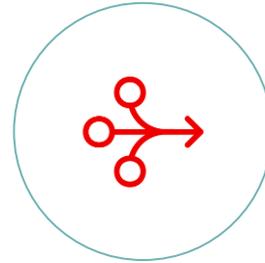# Dashboard user interface

Dashboard resources

The value of Red Hat OpenShift AI

What differentiates us?

Simplify AI adoption

Promotes freedom of choice and access to latest innovation on AI/ML technologies

Drive AI/ML operational consistency

Streamline the process of moving models from experiments to production

Gain hybrid cloud flexibility

Deploy models in containerized format across on-prem, clouds and edge, including disconnected environments

Red Hat Consulting AI Offerings

Learn how to maximize your technology investments

AI Accelerator

Rapidly deploy and adopt Red Hat OpenShift AI while advancing your AI practices:

● Upskill your ML Platform team and data scientists and adopt new AI capabilities
● Deploy an AI platform and tools with GitOps
● Review your current MLOps processes and identify improvements using Red Hat OpenShift AI

MLOps Foundation

Roll out automated MLOps pipelines and practices throughout your organization

● Establish self-service of your MLOps platform
● Automate and template ML pipelines
● Establish patterns and best practices for managing production ready solutions
● Transform models to microservices

AI Incubator

Upskill MLOps teams on AI tooling through delivery of gen AI use cases

● Work side-by-side with Red Hat experts utilizing best practices while learning MLOps and AI Engineering skills
● Prototype your custom gen AI solution
● Release a solution in your environment

Red Hat Learning: Developing and Deploying AI/ML Applications on Red Hat OpenShift AI (AI267)

Thank you

Red Hat is the world's leading provider of enterprise

open source software solutions. Award-winning

support, training, and consulting services make

Red Hat a trusted adviser to the Fortune 500.

linkedin.com/company/red-hat

youtube.com/user/RedHatVideos

facebook.com/redhatinc

twitter.com/RedHat

Red Hat