© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

May 2024

Let’s get real with AI

Ameeta Roy

Senior Director - Solution Architecture and Adoption -APAC

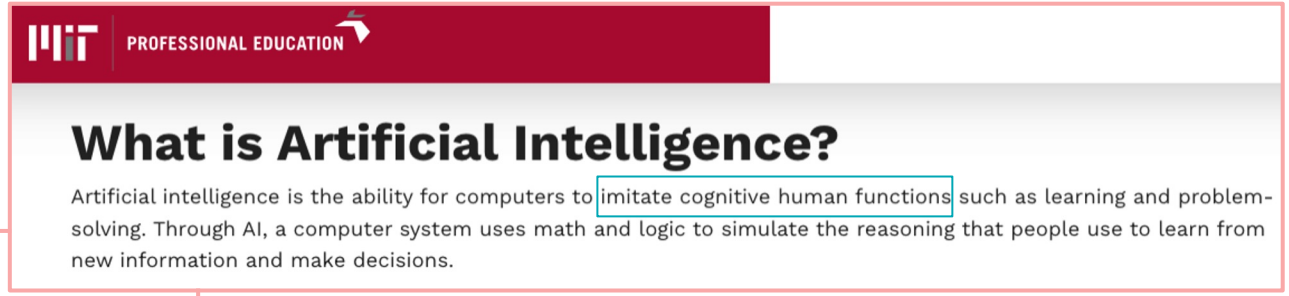
Red Hat

aws

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

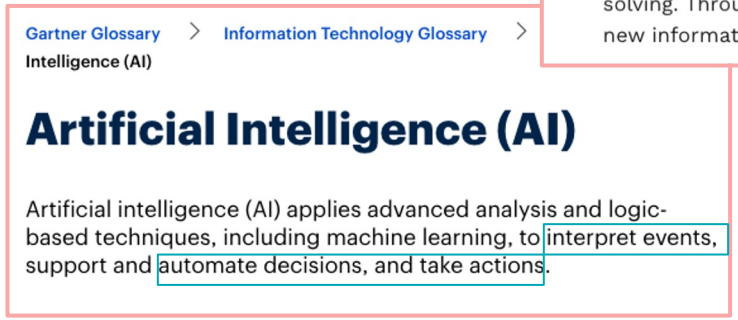
Source: https://www.redhat.com/en/blog/boston-university-sees-success-new-educational-tools-built-red-hat-openshift-data-science

What is artificial intelligence?


PROFESSIONAL EDUCATION

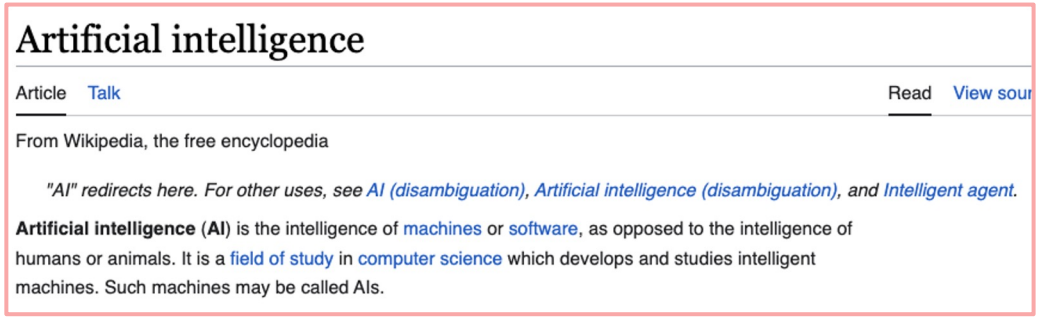
What is Artificial Intelligence?

Artificial intelligence is the ability for computers to imitate cognitive human functions such as learning and problem-solving. Through AI, a computer system uses math and logic to simulate the reasoning that people use to learn from new information and make decisions.


Gartner Glossary > Information Technology Glossary > Intelligence (AI)

Artificial Intelligence (AI)

Artificial intelligence (AI) applies advanced analysis and logic-based techniques, including machine learning, to interpret events, support and automate decisions, and take actions.


Artificial intelligence

Article Talk Read View sour

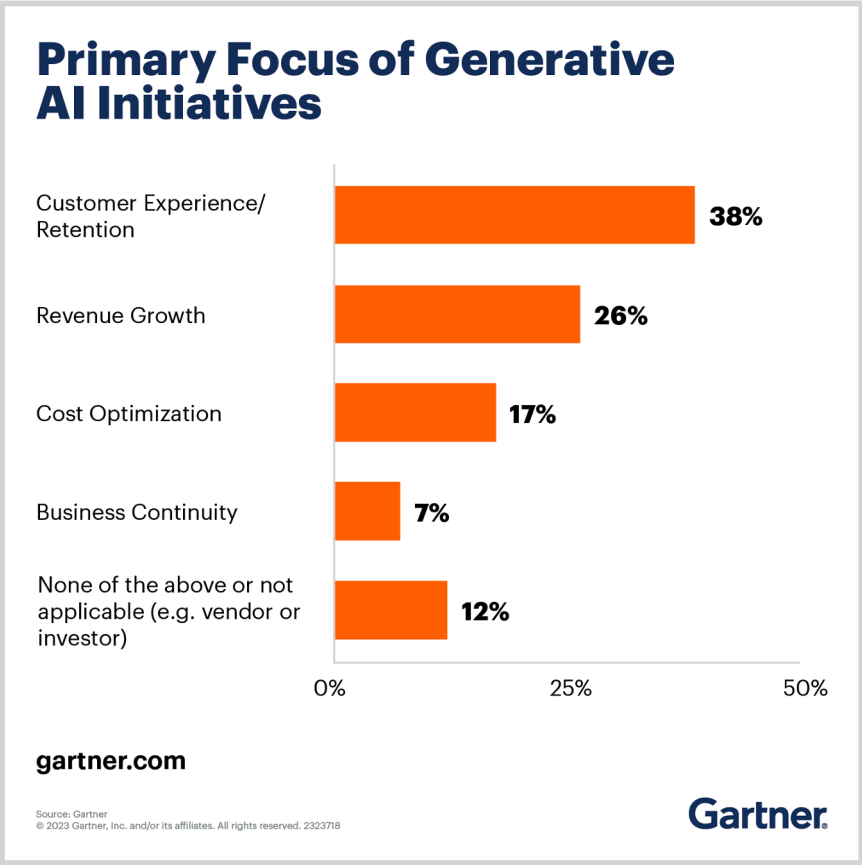
From Wikipedia, the free encyclopedia

"AI" redirects here. For other uses, see AI (disambiguation), Artificial intelligence (disambiguation), and Intelligent agent.

Artificial intelligence (**AI**) is the intelligence of machines or software, as opposed to the intelligence of humans or animals. It is a field of study in computer science which develops and studies intelligent machines. Such machines may be called AIs.

Red Hat

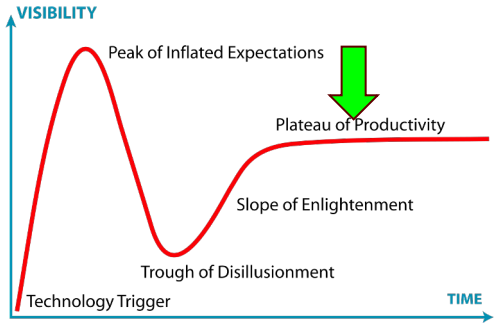
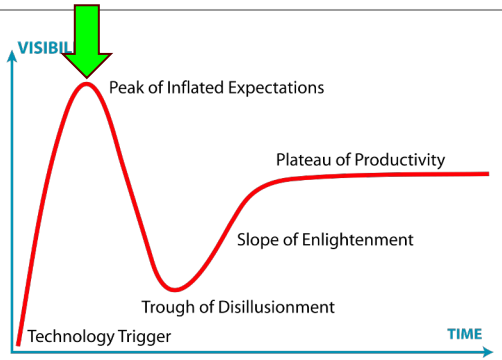
Generative AI: The Hype Is Real



- General-purpose technology
- Hype will lessen over time, as **the reality of implementation** sets in
- Impact will grow as people discover more innovative applications for the technology in their daily lives

Red Hat

Predictive AI vs. Generative AI

	Predictive AI	**Generative AI**
What is it for?	Predicts or forecasts outcomes based on historical data	Generates new, original content
Been around for	Decades	Years
Adoption	50% (plateaued)	33% (rapidly growing)
Maturity		
Situation	Implemented broadly across industries	Exploring use cases to provide value and differentiation

Source: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

Red Hat

Why we need to get this right

4 Common AI Project Challenges:

- Lack of Governance
- Ballooning costs
- Unrealistic expectations
- Data issues

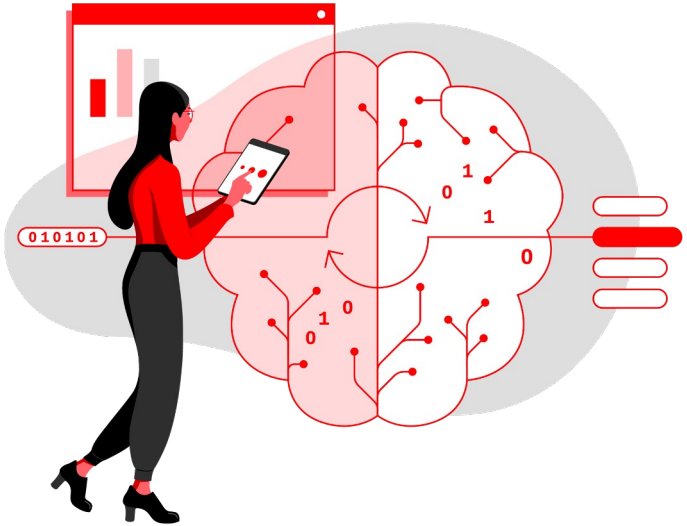
Source: https://www.cio.com/article/220445/6-reasons-why-ai-projects-fail.html

Red Hat

The Road to AI Success

Prioritise AI Use Cases

- Impact
- Risk
- Data



Source: https://www.cio.com/article/220445/6-reasons-why-ai-projects-fail.html

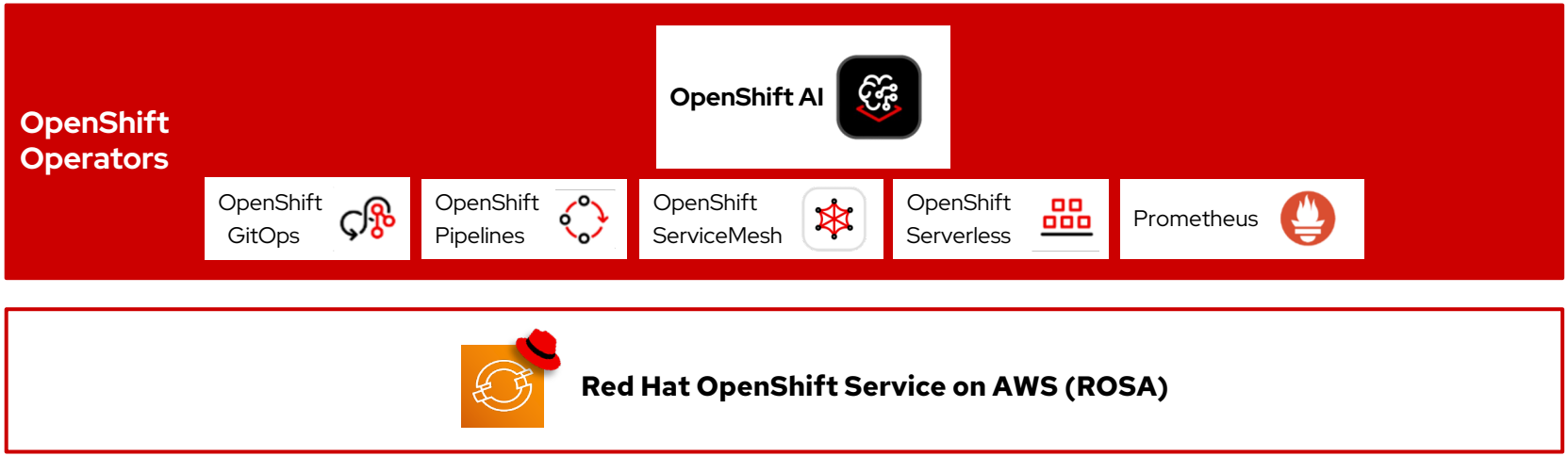
Red Hat

Where to now?

Image by storyset on Freepik

Red Hat

Red Hat OpenShift AI



OpenShift AI supports continuous development, integration, deployment and monitoring for AI/ML models

Bring your own model

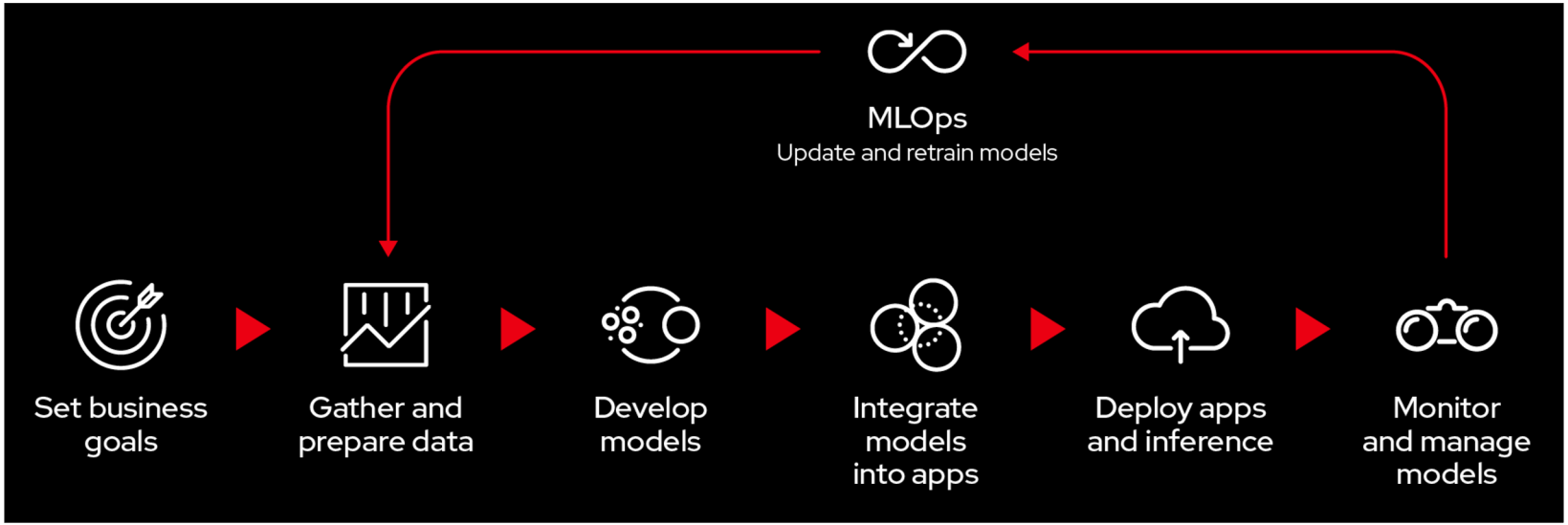
Best-of-breed AI technologies from a certified partner ecosystem

Red Hat provides the tools for
developing predictive models or **tuning** foundation models.

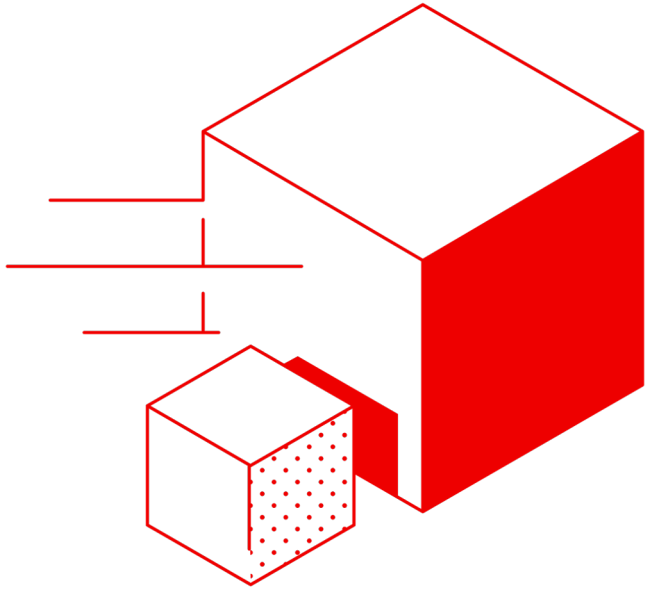
We don’t offer models, but partner with model builders and support the motion for ‘**bring your own model.**’

This is MLOps

Machine learning operations starts with understanding the use case.



Red Hat



Retrieval-augmented generation (RAG).

AI in 60 seconds *

Predictive

Predicts or forecasts outcomes based on historical data

Generative

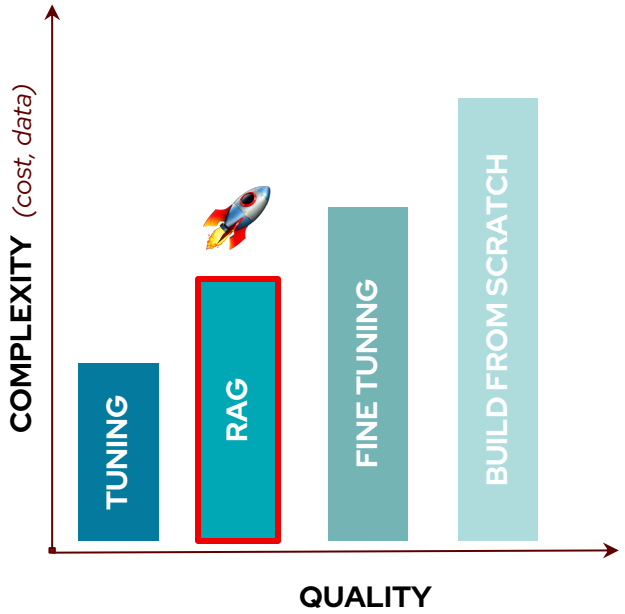
Uses larger set of data and models to generate new, original content

* we all wish!

Red Hat

Understanding foundational models for Generative AI

Where we are today

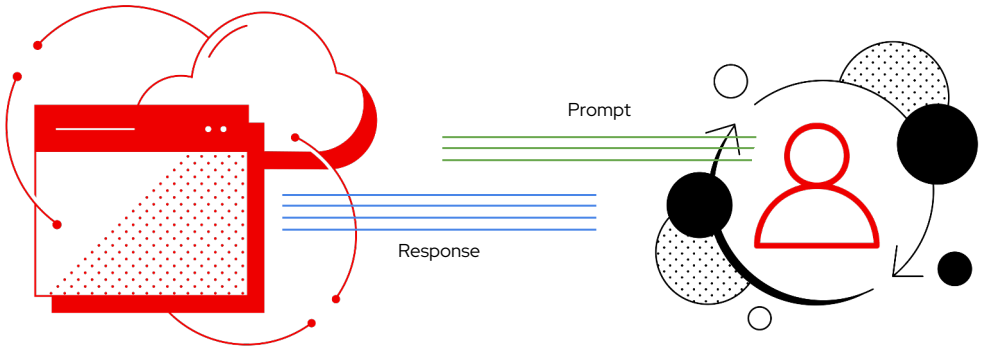


- **Prompt tuning:** Prepend prompt with data.
- **Retrieval augmented generation (RAG):** return facts from external sources.
- **Fine tuning:** Make small adjustments to the model data.
- **Training a foundation model from scratch:** can be a lot of work!

Red Hat

What is retrieval-augmented generation (RAG)?

LLM Only

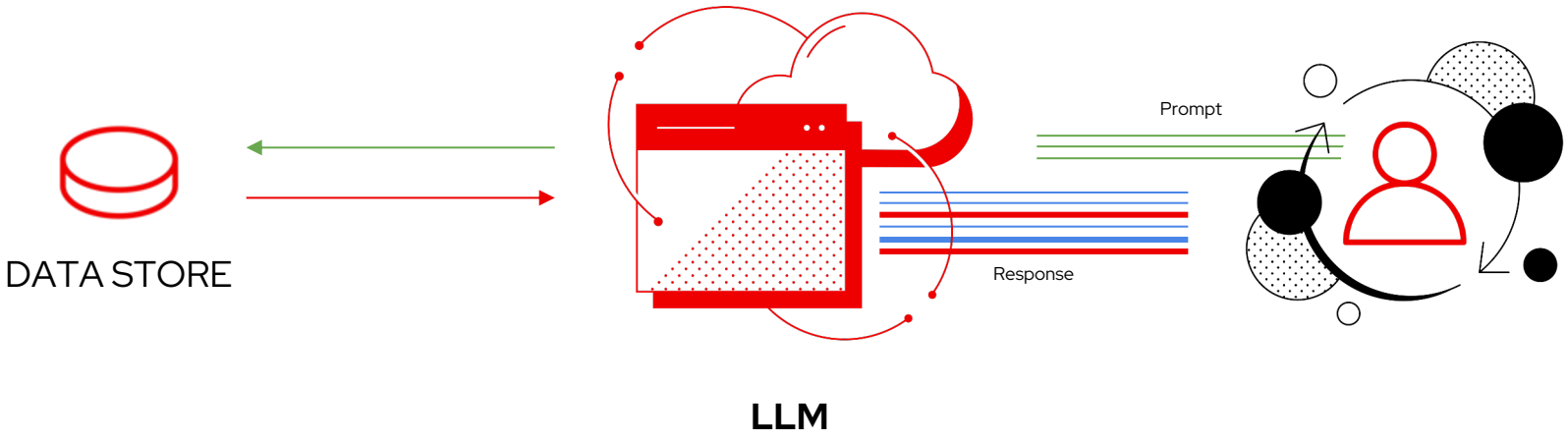


LLM

Red Hat

What is retrieval-augmented generation (RAG)?

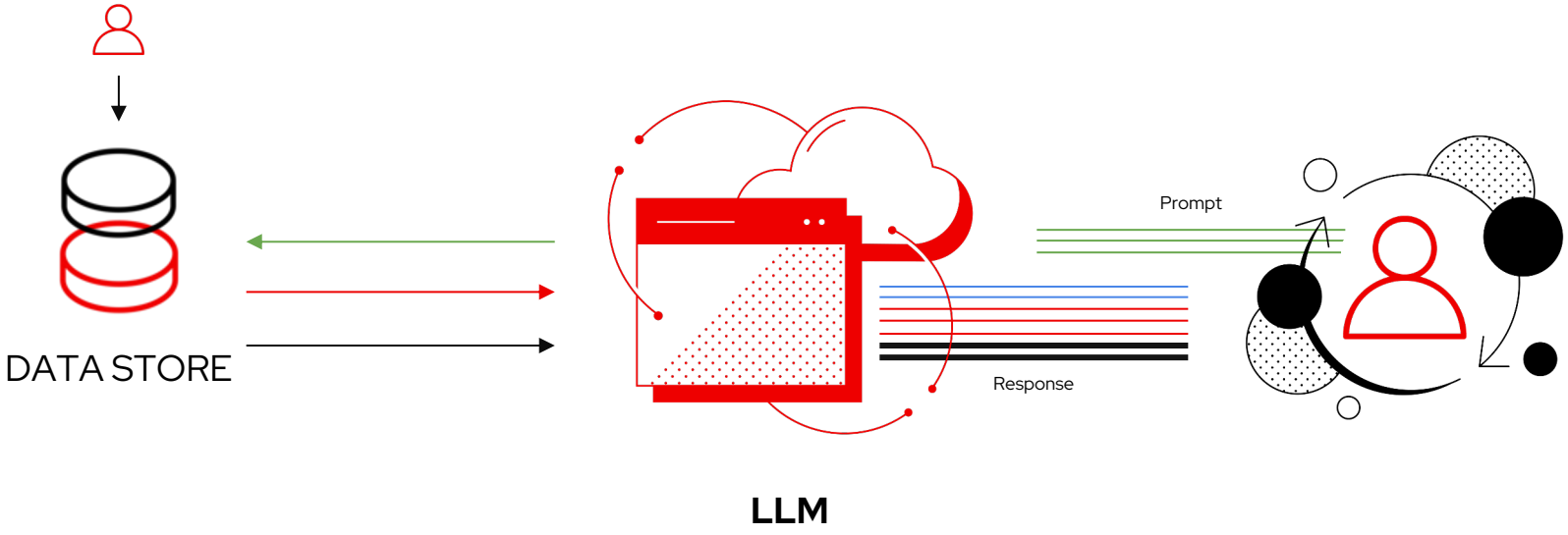
LLM + Data Store



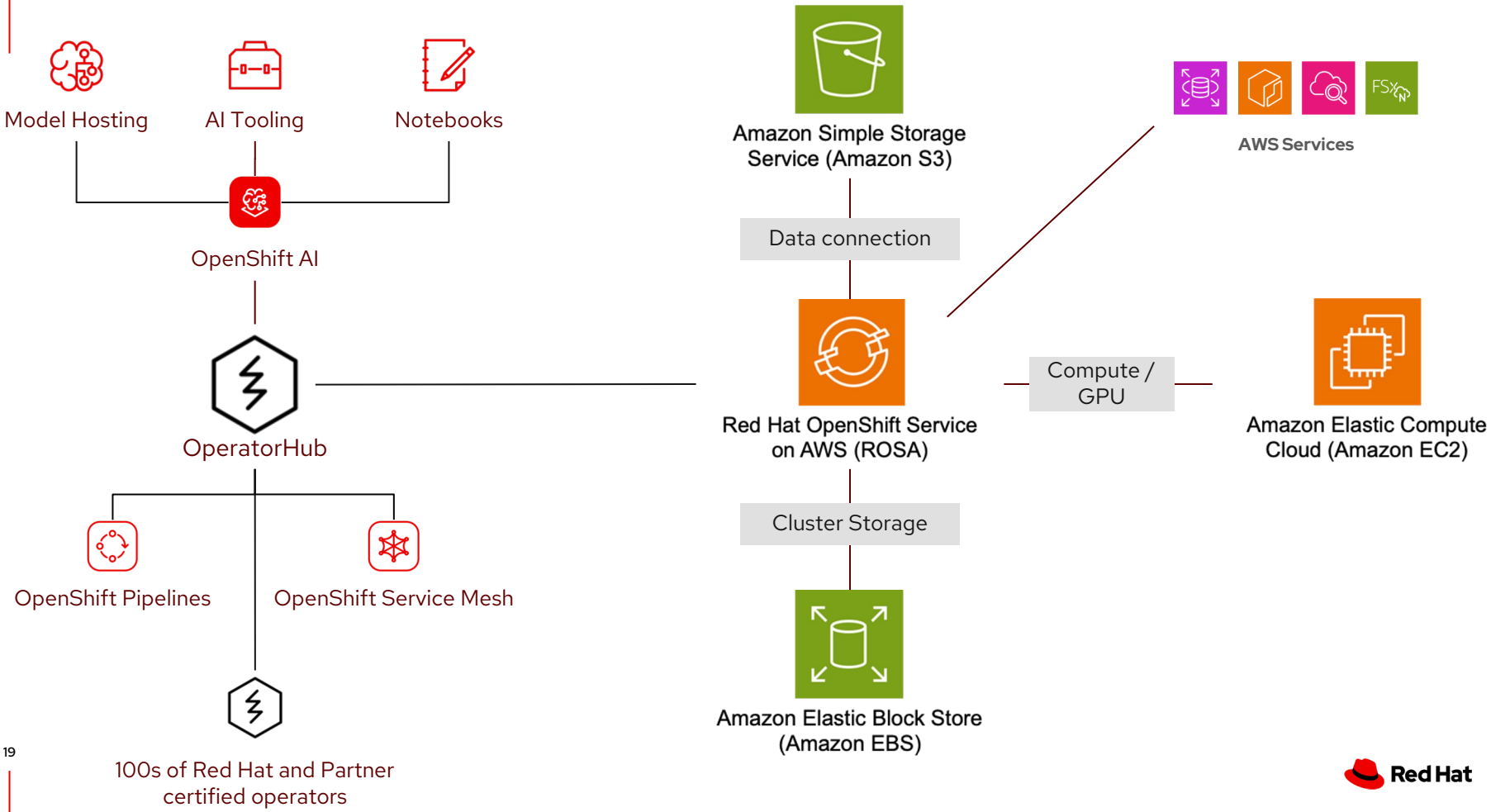
Red Hat

What is retrieval-augmented generation (RAG)?

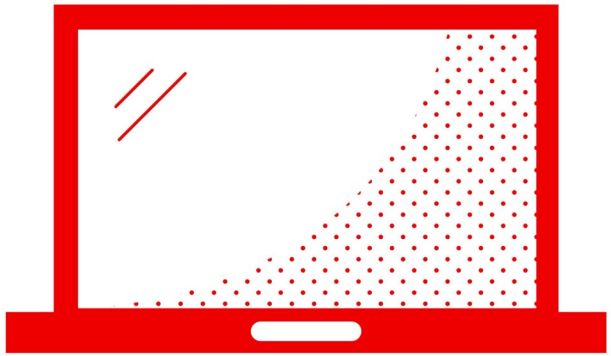
Augment data store with new data



Red Hat


Model Hosting
AI Tooling
Notebooks
OpenShift AI
OperatorHub
OpenShift Pipelines
OpenShift Service Mesh
100s of Red Hat and Partner certified operators
Amazon Simple Storage Service (Amazon S3)
Data connection
Red Hat OpenShift Service on AWS (ROSA)
Cluster Storage
Amazon Elastic Block Store (Amazon EBS)
AWS Services
Compute / GPU
Amazon Elastic Compute Cloud (Amazon EC2)

Red Hat



Let’s see this in action!

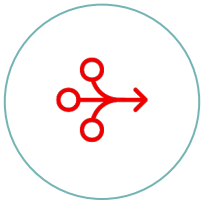
DEMO

The value of Red Hat OpenShift AI

A powerful platform for building AI/ML models and applications

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 to comply with regulatory requirements

Red Hat

Innovate while adding value

> The entire development and deployment process is vastly simpler with Red Hat OpenShift. We can deliver new solutions to the business and our customers faster than ever.
>
> **Charles Pizzato**
> Executive General Manager, IT Infrastructure, Suncorp Group

> *With Red Hat OpenShift Service on AWS, we have a service that is ready for the future. We don't have to worry about continuously improving the platform or managing the underlying infrastructure or operations, freeing our team to fully focus on providing greater innovation and value to our customers.*
>
> ARWIN SCHOLTEN, CHIEF TECHNOLOGY OFFICER, EDSN

Red Hat

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