

Accelerate artificial intelligence adoption in financial services



Global deployment of AI technologies is at an all-time high.

- ▶ 94% of business leaders believe that AI is critical to success over the next 5 years.²
- ▶ 79% of organizations deploy 3 or more types of AI applications.²
- ▶ 76% of leaders plan to increase their AI investments.²

Intelligent financial services applications support business success

Artificial intelligence (AI) and machine learning (ML) models use vast amounts of data to provide valuable insights, automate tasks, and improve core business capabilities. These technologies have the potential to transform all aspects of the financial services industry, from customers and employees to development and operations. Today's financial services organizations use AI/ML models to develop intelligent, cloud-native applications that deliver measurable outcomes like increased customer satisfaction, diversified service offerings, greater business and IT automation, and enhanced staff efficiency and productivity. In fact, 36% of financial services organizations report that AI/ML technologies reduce their annual costs by more than 10%.¹

Today, many financial services organizations deploy AI/ML models within business unit- or department-level solutions to enhance and augment activities currently performed by staff members. For example, ML algorithms can detect unusual transactions and spending patterns that human analysts might miss and flag them for investigation by fraud prevention teams. At the same time, AI/ML technologies can analyze account activity to identify suspicious patterns and recognize links between the seemingly unrelated accounts that are part of complex money laundering schemes. AI-powered chatbots can answer common support questions quickly and accurately—and even learn from previous interactions—to improve customer experiences and allow staff to focus on more complex issues. And by applying new ML algorithms to classic optical character recognition (OCR) technologies, financial institutions can improve the accuracy of scanned and digitized documents that are part of know your customer (KYC) processes.

However, building intelligent applications and deploying ML operations (MLOps) in production can be difficult. Financial services organizations must overcome several challenges to deliver innovative services based on AI/ML models:

- ▶ A limited number of available AI/ML experts makes it hard to find and retain data scientists and engineers, ML engineers, software developers, and other staff with the right knowledge.
- ▶ Lack of connections and collaboration between data scientists, ML engineers, developers, and other staff members slows AI/ML workflows and model life cycles.
- ▶ All phases of the AI/ML workflow—data preparation and management, model training and fine tuning, and inferencing—require expensive infrastructure that can be complicated to deploy and use efficiently in an automated, self-service manner.

¹ NVIDIA. "State of AI in financial services: 2023 trends," 2023.

² Deloitte. "State of AI in the enterprise, 5th Edition," October 2022.



Top AI use cases in financial services

Financial services organizations successfully use AI/ML for a variety of use cases:¹

- ▶ Natural language processing and large language models
- ▶ Recommender and next-best action systems
- ▶ Portfolio optimization
- ▶ Transaction and payment fraud detection
- ▶ Anti-money laundering and KYC initiatives
- ▶ Algorithmic trading
- ▶ Conversational AI systems
- ▶ Marketing optimization
- ▶ Risk analysis

- ▶ Training processes for AI/ML models must comply with data governance and sovereignty regulations that limit data sharing to certain geographies or departments within an institution.
- ▶ AI/ML solutions and pipelines are more complex than standard application deployments. They must support intensive model development, training, and inference processes; simple integration and deployment with intelligent applications; and the ability to monitor, update, and retrain models based on observed behaviors.

Overcome AI/ML challenges with hybrid cloud, container, and Kubernetes technologies

Intelligent, AI/ML-based applications require modern, cloud-native tools and best practices to manage development and deployment complexity. Containers and Kubernetes technologies provide the deployment agility, management capabilities, and scalability to deliver and manage intelligent applications. As lightweight, standalone units of software, containers package applications and dependencies—including runtimes, libraries, system tools, and settings—into easily distributable images. Kubernetes—an open source container orchestration platform—automatically creates, deploys, scales, and manages container instances across your environment.

Kubernetes-based container orchestration platforms offer several advantages as a foundation for AI/ML solutions. Automation and self-service capabilities let you provision MLOps environments on demand to speed and simplify AI/ML model development, testing, and deployment. Because containers are portable, you can use your models consistently across infrastructure footprints—including physical, virtualized, and private, public, and hybrid clouds—without change. Kubernetes automatically scales container workloads to ensure application availability while allocating resources only to workloads that need them. And, as a common technology platform, it lets you integrate components from a robust ecosystem of open source and commercial suppliers into your AI/ML solution with less effort.

Even so, effective AI/ML solutions need more than just standard Kubernetes functionality. Modern application platforms expand on Kubernetes to include advanced tools and greater capabilities. Continuous integration and deployment tools like Tekton and Jenkins can help you rapidly build, test, package, update, and deploy AI/ML models. GitOps continuous delivery tools—like ArgoCD—let you define and automate complex application deployments as code. Standardized monitoring services and application programming interfaces (APIs) aggregate information from all parts of your model in a single location so you can measure accuracy, detect bias, and initiate updates or retraining. Integration of containers and traditional applications that often run in virtual machines (VMs) lets you incorporate your existing tools into your AI/ML solution. And self-service connections to a broad range of container storage interface (CSI)-supported data stores and real-time data feeds simplify and improve model training.

Develop and deploy intelligent applications in a unified environment

Red Hat® technologies combine to create a unified, flexible MLOps foundation that supports the entire AI/ML life cycle, from model development and training to intelligent application integration and deployment. This solution provides the tools that you need to build and deploy innovative applications in a streamlined, consistent manner across all types of infrastructure, including physical hardware, virtual machines, and private, public, and hybrid clouds. A large selection of certified partner products and supported open source components lets you customize your environment to your organization's needs.



Financial services organizations run their AI/ML workloads in a variety of environments:

- ▶ 44% use hybrid cloud environments.¹
- ▶ 32% use only public cloud environments.¹
- ▶ 16% use only on-site datacenters.¹
- ▶ 5% use edge environments.¹
- ▶ 4% use private cloud environments.¹

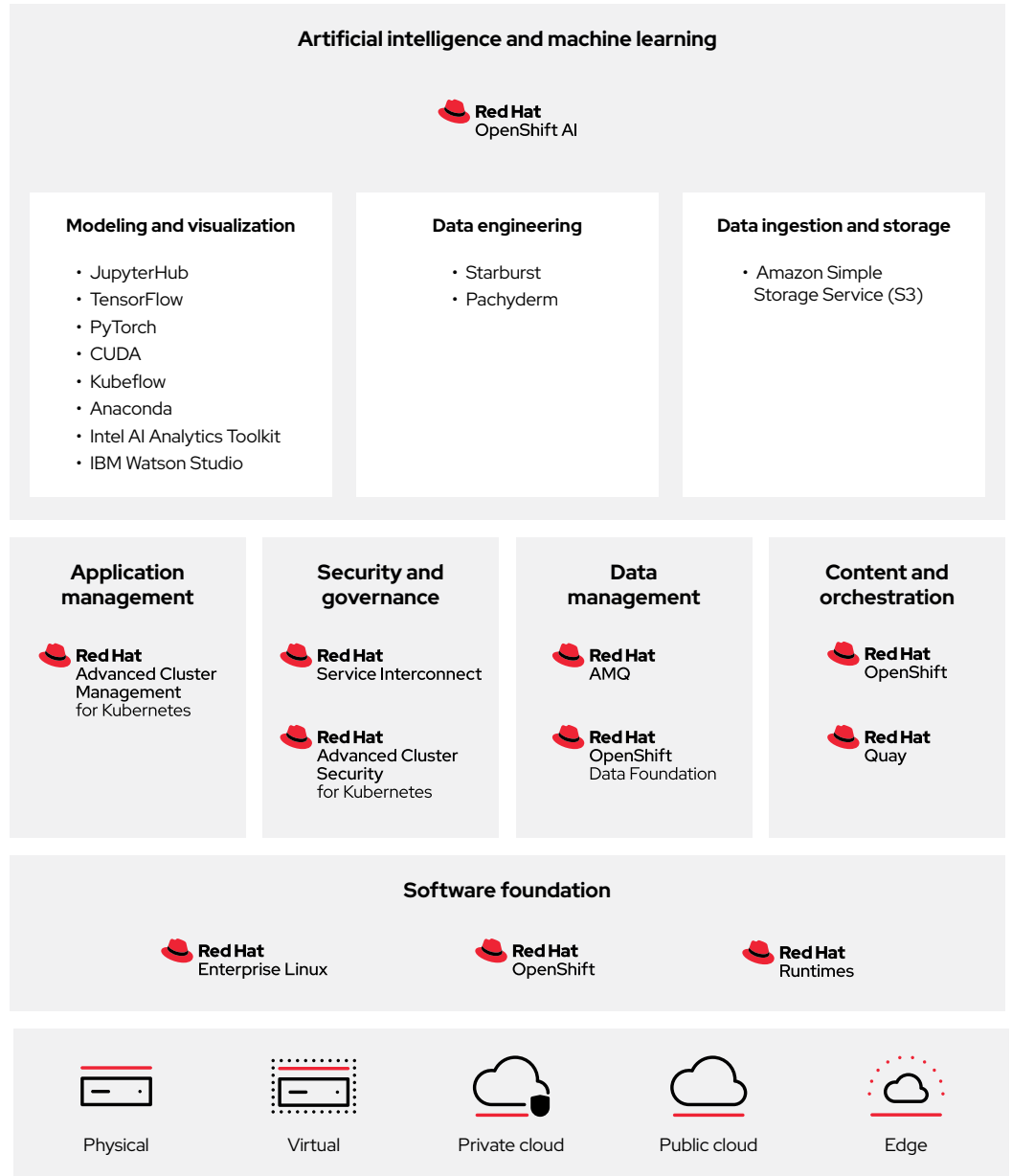


Figure 1. Red Hat brings together key technologies into unified solutions for consistent, standardized MLOps deployment.



Open source and AI/ML

Worldwide, researchers and developers use an open source development model to create many of the most advanced AI/ML technologies.

As the largest open source company in the world, Red Hat believes that open development models help create more stable, security-focused, and innovative technologies.

At Red Hat, we collaborate on community projects and protect open source licenses so we can continue to develop software that solves our customers' most complex challenges.

[Read more](#) about our commitment to open source.

Each layer in the solution performs key functions for developing, deploying, and managing AI/ML models and intelligent applications.

- ▶ The **software foundation** layer—including the operating system and container orchestration tools—is the consistent, stable, security-focused base for the entire AI/ML solution. Built-in [DevOps](#) and automation capabilities like self-service provisioning, along with a large set of included tools, languages, and runtimes, help you rapidly develop and maintain innovative, highly distributed, cloud-native AI/ML models and applications.
- ▶ The **application management** layer includes tools for deploying and managing cloud-native applications and services across hybrid cloud environments. Monitoring and policy enforcement features ensure that applications run consistently and efficiently.
- ▶ The **data management** layer consists of security-focused, scalable data streaming tools and structured and unstructured data storage for training and storing AI/ML models. Massively scalable storage lets you train the largest AI/ML models and maintain high accuracy rates.
- ▶ **Content and orchestration** tools and technologies help you manage your MLOps environment and workflows. Container registries and continuous integration and continuous deployment frameworks and workflows ensure that application build and deployment processes are consistent and repeatable across all infrastructure.
- ▶ The **security and governance** layer encompasses technologies to protect and prevent unauthorized access to AI/ML training and inference workloads. Increased security between datacenters and regions helps protect sensitive data in transit.
- ▶ The **artificial intelligence and machine learning** technologies layer includes certified tools and services for developing, training, and using AI/ML models. The underlying platform's self-service provisioning capabilities help you rapidly move from experiment to production in a collaborative, consistent manner.

Build an open, flexible MLOps foundation with Red Hat

As a leader in enterprise open source software, Red Hat provides a comprehensive product portfolio, proven expertise, and strategic partnerships with key independent software vendors (ISVs) to help you build your AI/ML solution architecture. Based on a set of curated, supported, and certified open source technologies, Red Hat delivers a security-focused foundation for building production-ready MLOps environments. Consistency across datacenter and cloud infrastructure lets you build AI/ML models and intelligent applications that deliver exceptional insight and user experiences.

Each component of the foundation delivers key functionality for MLOps.

[Red Hat OpenShift® AI](#) combines the proven capabilities of Red Hat OpenShift and Red Hat OpenShift Data Science into a consistent, scalable foundation for rapid application innovation and orchestrated deployment across infrastructures

- ▶ [Red Hat OpenShift](#) is a unified, enterprise-ready application platform for cloud-native development, deployment, and orchestration. On-demand compute resources and support for hardware and graphics processing unit (GPU) acceleration—including NVIDIA GPUs and Intel® Deep Learning Boost (Intel DL Boost)—speed modeling and inferencing tasks. Consistency across on-site, public cloud, and edge environments provides the speed and flexibility teams need to succeed. For exam-

**Customer success highlight:
Banco Galicia**

Working with Red Hat Consulting, Banco Galicia built an AI-based intelligent natural language processing (NLP) solution on Red Hat OpenShift, Red Hat Integration, and Red Hat's single sign-on technology.

Key outcomes:

- ▶ Reduced verification times from **days to minutes** with 90% accuracy
- ▶ Cut application downtime by 40%
- ▶ Improved agility by 4x

Read the [success story](#).

ple, you can create a self-service MLOps platform for data scientists, data engineers, and developers to rapidly build models, incorporate them into applications, and perform inferencing tasks. Collaboration features let teams create and share containerized modeling results in a consistent manner.

Red Hat OpenShift includes several key tools for intelligent application development and delivery. [Red Hat OpenShift Virtualization](#) lets you integrate, run, and manage existing virtualized applications as native Kubernetes objects. [Red Hat OpenShift Pipelines](#) lets you design AI/ML pipelines in a Kubernetes-native continuous integration and continuous deployment framework and run each pipeline step in isolated containers that scale on demand. And [Red Hat OpenShift GitOps](#) provides a continuous deployment workflow that automatically manages your MLOps environment based on configurations that you create and store in Git.

Included with Red Hat OpenShift, [Red Hat Enterprise Linux](#)[®] provides a consistent, scalable, and high-performance operating foundation across datacenter, cloud, and edge environments. Built-in security features like [Security-Enhanced Linux \(SELinux\)](#) defend against threats and help you stay in compliance with industry and regulatory environments. And because Red Hat platforms work together, these features and certifications extend throughout the Red Hat software stack.

- ▶ [Red Hat OpenShift Data Science](#) is an AI platform offering based on the open source [Open Data Hub](#) project. It gives data scientists and developers a powerful AI/ML platform for gathering insights and building intelligent applications. It includes tools and frameworks like Jupyter, TensorFlow, and Pytorch. The service also incorporates components from technology partners like Starburst, Anaconda, IBM, and Intel to further speed development of intelligent financial services applications.

[Red Hat Integration](#) is a comprehensive set of integration and messaging technologies that connect applications and data across hybrid infrastructures. Key components within Red Hat Integration:

- ▶ [Red Hat Runtimes](#) is a set of products, tools, and components—including lightweight runtimes and frameworks—that help you develop and maintain highly distributed, cloud-native, AI/ML applications.
- ▶ [Red Hat AMQ](#) is a scalable, flexible, distributed messaging platform that streams data to your AI/ML models with high throughput and low latency.

[Red Hat OpenShift Data Foundation](#) is software-defined storage that provides scalable persistent file, block, and object storage for the largest AI/ML datasets.

[Red Hat Service Interconnect](#) is an over-the-top communications protocol that simplifies connectivity between applications and services that span multiple datacenters or regions. Anyone on your development team can use the protocol without needing elevated privileges or compromising security.

[Red Hat Advanced Cluster Management for Kubernetes](#) is a unified console for controlling, automating, and monitoring application deployment, cluster management, and policy enforcement at scale and based on geography across your cloud environment.

[Red Hat Advanced Cluster Security for Kubernetes](#) helps protect containerized Kubernetes workloads across hybrid cloud deployments to enhance the security of your applications.

[Red Hat Quay](#) is a security-enhanced private registry platform for managing cloud-native content across hybrid cloud environments.

Finally, our [certified partner ecosystem](#) lets you integrate popular AI/ML, data analysis, management, storage, security, and development tools into this architecture. We work closely with partners to certify



“Red Hat plays a role in implementing NLP within Banco Galicia by providing us with the technology and architecture. Through Red Hat, we managed to understand everything to do with Red Hat OpenShift. We also started to design an architecture that could be cloud-native.”

Matias Lorusso
Solution Architect,
Banco Galicia

their software on our platforms for increased manageability, security, and support. Many partners also provide certified [Red Hat OpenShift operators](#) to simplify software life cycle management.

Take advantage of large language models

Within the financial services industry, document digitization, report analysis, and conversational service applications often take advantage of large language models (LLMs) like GPT-4 (Generative Pre-trained Transformer 4). Due to the effort and compute resources required to create LLMs, organizations commonly use pretrained models in these applications. Even so, these models still need additional domain- or company-specific training and fine-tuning with a smaller set of local data to deliver accurate results. And each application requires its own specifically trained and tuned model.

Red Hat OpenShift AI is an ideal platform for training and tuning LLMs. A single, scalable, consistent platform for model development, training, and inferencing and application integration and deployment eliminates duplicate work and improves the efficiency of the entire AI/ML life cycle. Advanced cluster management and self-service capabilities let you build AI/ML pipelines that are reusable and repeatable across multiple models and applications. In fact, organizations that deployed Red Hat OpenShift AI experienced a 20% improvement in data scientist efficiency.³

Learn more

Red Hat provides a complete technology portfolio, proven expertise, and strategic partnerships to help you achieve your goals. We deliver a foundation for building production-ready AI/ML environments in addition to services and training for rapid adoption. No matter where you are in your AI/ML journey, we can help you build a production-ready MLOps environment that accelerates development and delivery of intelligent applications.

- ▶ [Learn more](#) about AI/ML solutions for the financial services industry.
- ▶ [Schedule a complimentary discovery session](#) to find out how we can help you deploy a foundation for MLOps.

³ Forrester Consulting study, commissioned by Red Hat. “[The Total Economic Impact™ of Red Hat Hybrid Cloud Platform for MLOps](#),” March 2022. Results are for a composite organization representative of interviewed customers.



About Red Hat

Red Hat is the world’s leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers develop cloud-native applications, integrate existing and new IT applications, and automate and manage complex environments. [A trusted adviser to the Fortune 500](#), Red Hat provides [award-winning](#) support, training, and consulting services that bring the benefits of open innovation to any industry. Red Hat is a connective hub in a global network of enterprises, partners, and communities, helping organizations grow, transform, and prepare for the digital future.

North America

1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa

00800 7334 2835
europe@redhat.com

Asia Pacific

+65 6490 4200
apac@redhat.com

Latin America

+54 11 4329 7300
info-latam@redhat.com

[facebook.com/redhatinc](#)
[@RedHat](#)
[linkedin.com/company/red-hat](#)

redhat.com
483536_0823_KVM