Why Open Source Artificial Intelligence Platforms Help Enterprise Business Transformation

Ritu Jyoti
Group Vice President, Worldwide Artificial Intelligence and
Automation Research Practice, Global AI Research Lead, IDC

Table of Contents

CLICK ANY HEADING TO NAVIGATE DIRECTLY TO THAT PAGE.

Introduction

Artificial intelligence (AI) is the most disruptive innovation of our lifetime. As per IDC's *AI StrategiesView 2022* study — a global survey of 2,053 organizations, with IT and line-of-business decision makers and influencers as respondents — AI adoption and spend is on the rise with large enterprises spending an average of $200 million annually. Early adopters report about 35% improvement in innovation and about 32% improvement in employee and customer experiences with the rollout of AI solutions.

Enterprises across industries are embracing AI/machine learning (ML) for a breadth of use cases. For example, AI/ML in financial services is helping to improve loan underwriting and reduce risk. AI can also lessen financial crime through advanced fraud detection and spotting anomalous activity. Deep learning (DL) algorithms are being used to shave down the time it takes to diagnose serious illnesses. Machinery maintenance and quality are the leading AI transformation projects in manufacturing operations today. AI is making predictive maintenance a reality for industrial IoT.

Most AI systems so far have been used for classification and predictions. The rise of generative AI has the potential to be a major game changer for businesses due to its ability to be creative. This technology is a branch of computer science that involves unsupervised and semi-supervised algorithms and enables computers to create new content using previously created content, such as text, audio, video, images, and code in response to short prompts.

The application of generative AI in enterprises is just starting to unfold along with a world of opportunity. Enterprises can leverage generative AI to drive innovation, automate repetitive tasks, improve decision making, personalize customer experiences, and boost efficiencies. As such, businesses that can effectively leverage technology are likely to gain a significant competitive advantage.

As advancements in AI/ML rise across the world, business leaders are embracing the open source culture and technologies. There are many reasons for this, including multi-organizational collaboration, accelerated innovation, and social responsibility. A software-maker culture inspires established companies to experiment with AI and create platforms that align with their business needs. These companies already have the data that fuels AI; they need the freedom to build and deploy models as rapidly as the business context changes. Using an organization's data to train these models gives companies the domain-specific context to make better business predictions and decisions. Open source helps them become co-inventors of software, not just buyers of it. Open source AI has been one of the most important developments within the technology industry, and it is continuing to grow.

Organizations typically embrace one or more of these options to support their AI/ML initiative life cycle: fully managed service, self-managed service, or commercial offerings they can customize to their needs. As per IDC's *ML and Cloud Survey* in North America, with a sample size of 400, about one-third of participants reported using open source DIY or self-managed service, and two-thirds reported a combination of various options.

However, AI/ML initiatives are not without challenges and risks. As per IDC's *AI StrategiesView 2022,* cost, lack of machine learning operations (MLOps) tools (covered in detail in the later sections of this paper) and lack of responsible AI tools are the top 3 challenges. Interestingly, downloading open source software may introduce the risk of lack of governance and security. It may spin up shadow IT and drive associated risks. But when embraced appropriately with the right security and management layers, it can drive significant business benefits of flexibility, agility, portability, and overall time to value as noted from the customer interviews and their success stories captured in a section later in this paper.
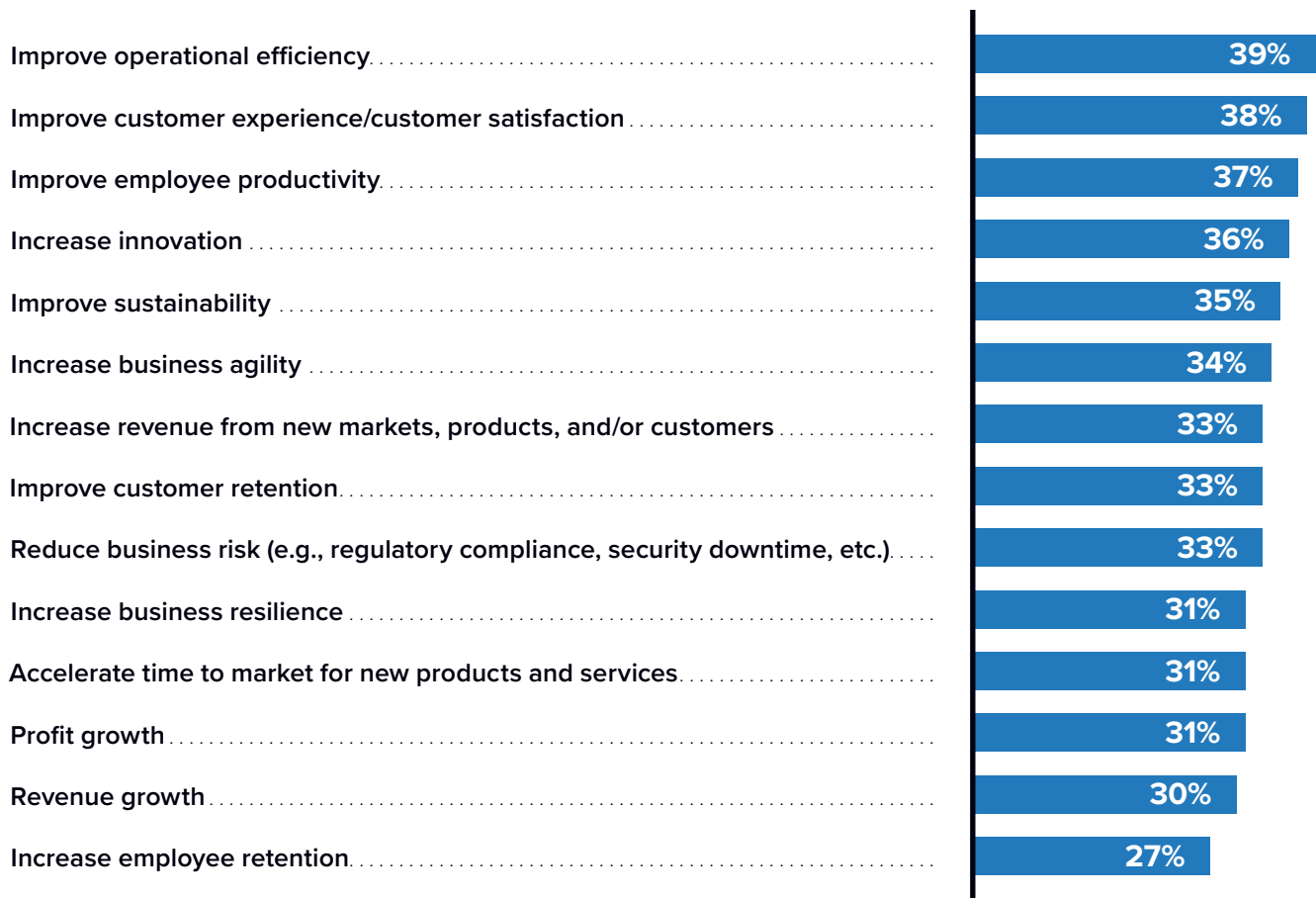
AI/ML Adoption Trends and Challenges

As per IDC's *AI StrategiesView 2022,* two of the leading drivers for AI adoption are improving operational efficiency and improving customer experience and satisfaction (see **Figure 1**).

FIGURE 1

Primary Business Drivers for AI Adoption

What are the primary business objectives for using AI for your projects?
(% of respondents)

Improve operational efficiency	39%
Improve customer experience/customer satisfaction	38%
Improve employee productivity	37%
Increase innovation	36%
Improve sustainability	35%
Increase business agility	34%
Increase revenue from new markets, products, and/or customers	33%
Improve customer retention	33%
Reduce business risk (e.g., regulatory compliance, security downtime, etc.)	33%
Increase business resilience	31%
Accelerate time to market for new products and services	31%
Profit growth	31%
Revenue growth	30%
Increase employee retention	27%

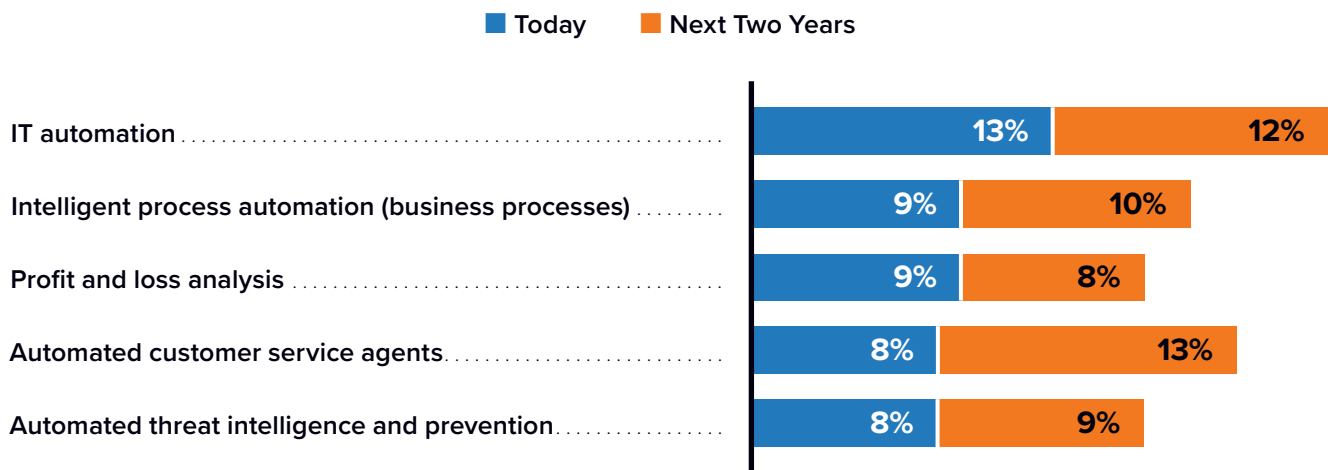n = 2,053; Source: IDC's *AI StrategiesView 2022,* CY22

This is reflected in the leading use cases for AI, which include automated customer service agents, sales process recommendation and automation, automated threat intelligence and prevention, and IT automation. Combined, these four use cases will represent nearly a third of all AI spending this year. Some of the fastest-growing use cases are automated human resources, fraud detection, risk analytics, and predictive maintenance.

AIOps or IT automation is the number 1 use case for using AI/ML (see **Figure 2**). It has strong interdependencies with MLOps.

FIGURE 2

AI Solutions: Leading Use Cases

For what use cases are you currently deploying, building, or investigating AI applications/solutions?
(% of respondents)

■ **Today**   ■ **Next Two Years**

Use Case	Today	Next Two Years
IT automation	13%	12%
Intelligent process automation (business processes)	9%	10%
Profit and loss analysis	9%	8%
Automated customer service agents	8%	13%
Automated threat intelligence and prevention	8%	9%

n = 2,053; Source: IDC's *AI StrategiesView 2022,* CY22
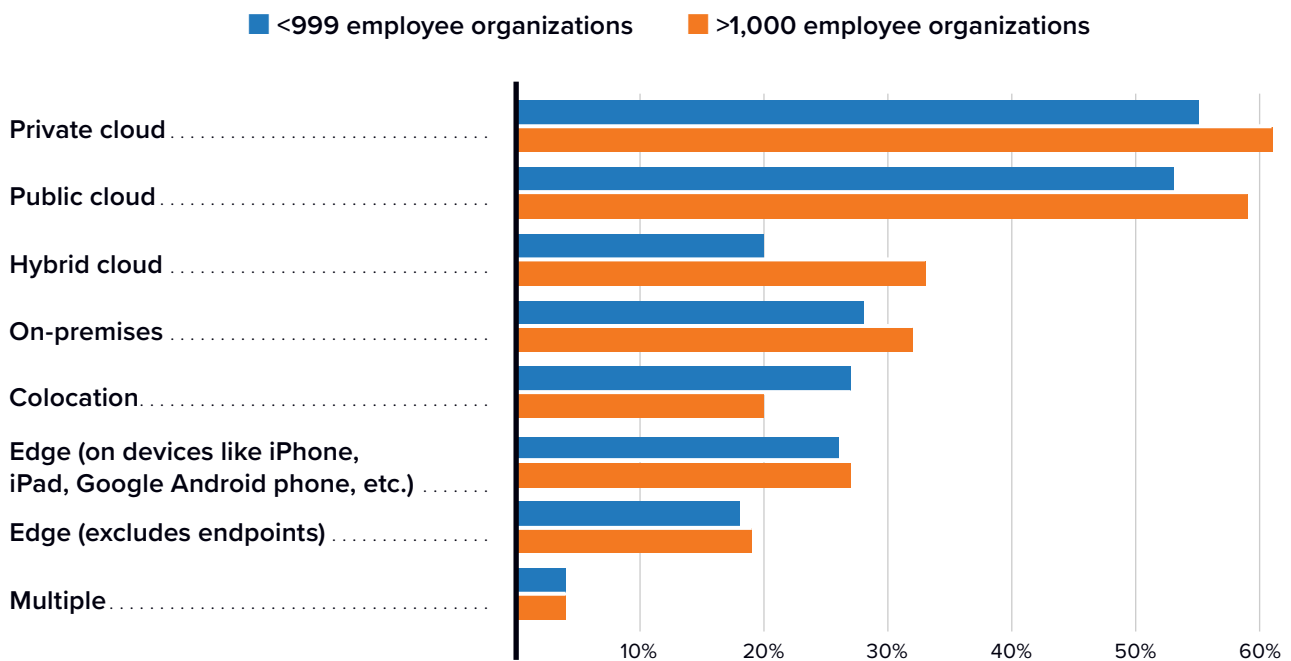
For an accessible version of the data in this figure, see Figure 2 Supplemental Data in the Appendix.

AIOps or IT automation is the number 1 use case for using AI/ML. It has strong interdependencies with MLOps.

Deployment locations for AI/ML solutions are distributed (see **Figure 3**). Private cloud and public cloud are the largest percentage for deployment location for AI/ML solutions. Hybrid cloud is getting close to traditional on-premises. Colocation and edge are close for large firms and growing rapidly. Multicloud is still nascent.

FIGURE 3

AI/ML Solutions: Deployment Locations

Where are your organization's AI applications/solutions deployed?
(% of respondents)

■ **<999 employee organizations**　　　■ **>1,000 employee organizations**

n = 2,053 (total), n = 348 (<999), n = 1,705 (>1,000); Source: IDC's *AI StrategiesView 2022,* CY22

For an accessible version of the data in this figure, see Figure 3 Supplemental Data in the Appendix.
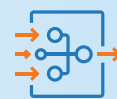
These four use cases will represent nearly a third of all AI spending this year:

Automated customer service agents

Sales process recommendation and automation

Automated threat intelligence and prevention

IT automation

As per the same IDC's *AI StrategiesView 2022* study, the top challenges (see **Figure 4**) in implementing AI/MI solutions range from the cost of the solution (which includes the cost of underlying infrastructure and tools) to a lack of machine learning operations tools and a lack of responsible AI tools.

**FIGURE 4**

Top Challenges in Implementing AI/ML Solutions

What are your top challenges in implementing AI/ML solutions?
(% of respondents)

55%	52%	42%
Cost	Lack of MLOps tools	Lack of responsible AI tools
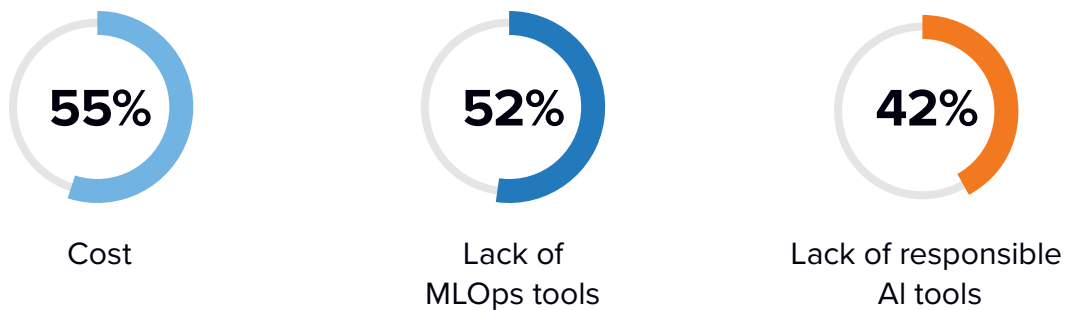
n = 2,053; Source: IDC's *AI StrategiesView 2022,* CY22

While research in AI is making breathtaking advances, large enterprises still struggle to apply AI/ML technologies successfully. As noted above, lack of MLOps tools is a critical challenge. In the sections that follow, let us understand what MLOps is and examine the key requirements for efficient operations and how open source AI and a modern application platform can help alleviate the associated challenges. In addition, let us also examine the synergies between hybrid cloud and AI and how open source AI can help. Lastly, let us do a quick assessment of how open source AI can help mitigate issues related to cost.
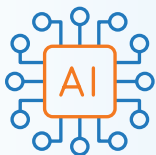
> While research in AI is making breathtaking advances, large enterprises still struggle to apply AI/ML technologies successfully.
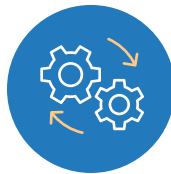
Machine Learning Operations (MLOps)

MLOps — a compound of machine learning, development, and operations — is the practice of collaboration between data scientists, business analysts, data architects, ML engineers, and operations professionals to help manage production AI/ML life cycles. It is powered by tools and technologies to support each stage of the life cycle, from experimentation to production, and in a continuous feedback loop to support data and concept drift.
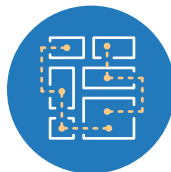
Requirements

It should integrate or align with existing DevOps practices, while also delivering the additional unique capabilities required to manage ML deployment.

The MLOps toolchain needs to provide visibility, managed access control, and collaboration for all of the practitioners involved in the machine learning life cycle.
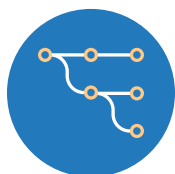
Data scientists, ML engineers, and application developers need access to their preferred tools and resources to be most productive.

IT operations teams need to ensure that resources are up to date, in compliance, and used in a secure manner.

Data scientists and ML practitioners have many of the same concerns that software engineers do: repeatable experiments (like repeatable builds); portable and reproducible environments (like having identical setups in development, stage, and production); credential management; tracking and monitoring metrics in production; flexible routing; and effortless scale out.

Enterprise AI/ML initiatives require compatible tools and simple management of the evolving software stack.

How Open Source AI/MLOps Platforms Help

Containers let data science and ML practitioners quickly and easily deploy a broad selection of AI/ML tools across hybrid environments in a consistent and reproducible fashion. Practitioners can iteratively modify and share container images with versioning capabilities that track changes for transparency, while process isolation and resource control improve protection from threats.

Kubernetes, an orchestration tool of choice by customers, lets practitioners administer the life cycle of the containers in an efficient way, typically centralizing access to compute, storage, and networking resources across onsite, edge, data, and cloud environments. Kubernetes also provides unified workload scheduling, multitenancy controls, and quota enforcement that run within containers. Data scientists, ML engineers, and application developers can deploy tools and applications from a pre-approved service catalog — without needing to engage with IT operations teams.

Automation also frees staff from tedious tasks, allowing them to focus on more strategic activities. IT operations teams can automate common life-cycle management tasks such as configuration, provisioning, and updates to gain efficiency, speed, and accuracy.

The Kubeflow project, targeted at ML engineers who need to stand up and maintain machine learning workloads and pipelines on Kubernetes, provides templates and custom resources to deploy a range of machine learning libraries and tools on Kubernetes. Kubeflow is an excellent way to run frameworks like TensorFlow, JupyterHub, Seldon, and PyTorch under Kubernetes and thus represents a path to truly portable workloads on Kubernetes: A data scientist or ML engineer can develop a pipeline on a laptop and deploy it anywhere.

Best Practices and Recommendations

Look for a robust, highly available application platform built on containers that includes integrated security features and makes it easy to deploy, manage, and move containers across your environment. Choose an open source platform that integrates with a broad set of technologies to gain more flexibility and choice. Select a full-featured Kubernetes distribution — including continuous integration/continuous delivery (CI/CD) capabilities and tooling rolled in — to take advantage of a leading open source technology.

Choose container application life-cycle management tools that include easy-to-use automation and integration with your preferred AI/ML tools. Popular options include Kubernetes Operators and Helm Charts.

Embrace MLOps approaches that foster collaboration between AI/ML teams, application developers, and IT operations to accelerate time to production for ML-powered applications. Automation — often in the form of CI/CD pipelines — makes rapid, incremental, and iterative change possible for faster application development life cycles.

MLOps is more than just technology — it involves people and processes and supports the open source culture of being a co-innovator. Apply MLOps approaches to the entire AI/ML life cycle. Take advantage of automation in your platforms and tools, as well as open source technologies like Argo CD, Tekton, Jenkins, and Spinnaker, to create CI/CD pipelines.

Synergies Between Hybrid Cloud and AI Initiatives and the Value of Open Source AI

Data fuels AI, and the quality of the data directly affects the results of AI. The cloud and mobile revolutions have accelerated the pace of data creation, both structured and unstructured. Hybrid platforms are the future of data architecture and address the data proliferation challenge across multiple silos, databases, and clouds. They knock down data silos to bring AI to their data. AI and hybrid architectures are enhanced by the extraordinary synergy between the two.

One important consideration for organizations developing AI solutions using proprietary public cloud fully managed services is the limitation of those services in accessing critical on-premises data because of security, data gravity, and compliance issues which would skew data resulting in less than optimal business outcomes. No single algorithm or solution can solve a complex business problem. This represents modern-day "vendor lock-in," which can stifle innovation and prevent businesses from scaling AI efforts and using various open source products and services. AI and data tools that can operate on a hybrid multicloud platform allow organizations to take advantage of their data and applications across any cloud (public or private) and on premises via containers in a consistent way. As enterprises modernize for an AI and hybrid cloud world, they will find there is less "assembly required" in expanding the impact of AI across the organization.

Open source platforms and cloud-native architecture accelerate development, enable portability and access, and simplify interoperability between clouds. One common struggle is access to what is needed; if it is defined as part of container authorization, it eases that challenge. Those operational advantages are also applicable in AI and machine learning development.

Hybrid platforms are the future of data architecture and address the data proliferation challenge across multiple silos, databases, and clouds. They knock down data silos to bring AI to their data. AI and hybrid architectures are enhanced by the extraordinary synergy between the two.

Hybrid cloud platforms thus help operationalize and scale AI through modernization of applications and infrastructure. With high-end CPUs or GPU farms, it is far easier and cheaper to access them on demand in a cloud and utilize a pay-as-you-go model. To harness all of the compute horsepower, AI programmers need a special set of software frameworks (such as TensorFlow, MXNet, Caffe, Watson, and PyTorch) and programming languages, both general and specialized (such as Python, Java, R, and Prolog). These requirements are better met in a hybrid multicloud environment.

Other considerations for AI applications are high-performance storage systems and high-bandwidth networks, both of which can be effectively addressed by a well-designed hybrid multicloud unified architecture. The underlying Kubernetes infrastructure should also provide some level of consistency, whether running in the public cloud or on premises.

Considering Red Hat OpenShift AI as the Hybrid MLOps Platform

Now, let us do a high-level review of Red Hat OpenShift AI and examine how it helps support the machine learning operations requirements and challenges noted above.

OpenShift AI is available as a fully managed cloud service or traditional software that can run on-premises, in the cloud, or at the edge (see **Figure 5**, next page). It is an add-on to Red Hat OpenShift, an enterprise-grade open source application platform for accelerating the development and delivery of cloud-native applications.

Other considerations for AI applications are high-performance storage systems and high-bandwidth networks, both of which can be effectively addressed by a well-designed hybrid multicloud unified architecture.

OpenShift AI originated from Open Data Hub, an open source community project that serves as a blueprint for building an AI-as-a-service platform on Red Hat OpenShift. Open Data Hub inherits from upstream open source projects such as Kafka/Strimzi and Kubeflow and is the foundation for Red Hat's own internal data science and AI platform.

With OpenShift AI, data scientists can create models using Jupyter notebooks and select from popular tools such as TensorFlow, scikit-learn, PyTorch and more for developing models. Components of Kubeflow are incorporated for managing and automating the machine learning lifecycle, with a focus on MLOps.

Red Hat OpenShift AI has a robust partner ecosystem with optional integrated ISV solutions to complement core platform capabilities and access to the full partner ecosystem with over 30 AI/ML technology partners and their associated solutions.

FIGURE 5

OpenShift AI: AI Platform for Hybrid Cloud
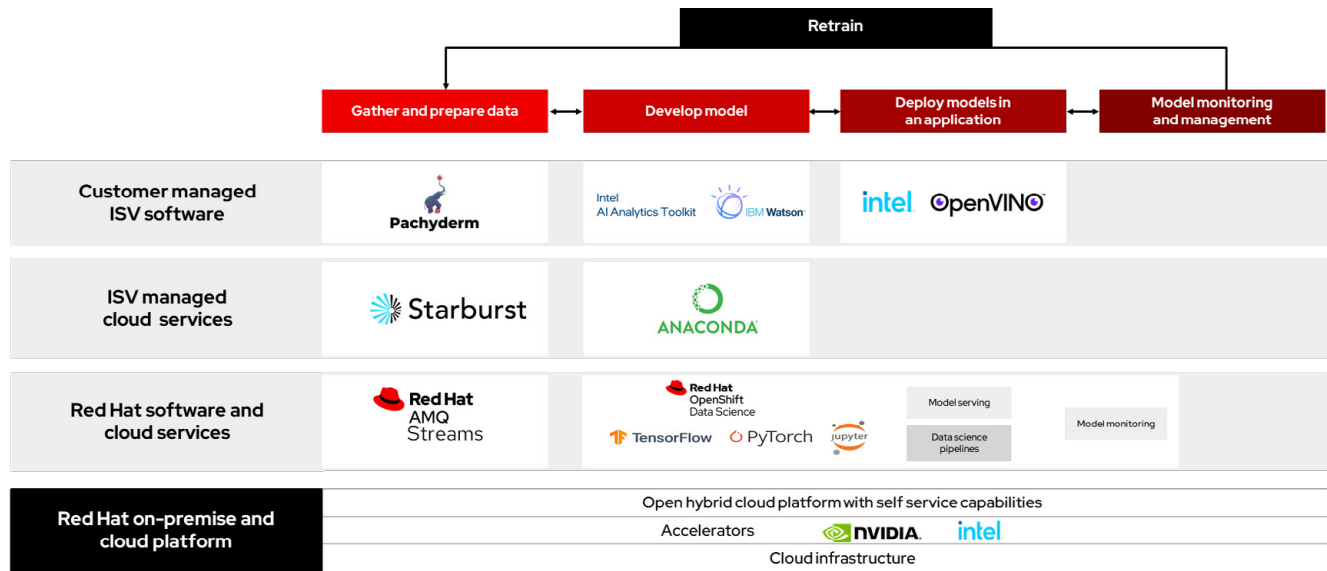
Source: RedHat, 2023

Customer Scenarios

As part of this research, IDC interviewed multiple Red Hat customers who have been successfully using Red Hat OpenShift AI for accelerating the AI/ML life cycle.

The following outlines three example customer scenarios and success stories.

Risk management and improvement in profits:

A large commercial bank in Turkey initiated modernization of its on-premises infrastructure a couple of years back. Due to regulations, it operates 100% on premises. OpenShift adoption is its steppingstone to embracing more cloud-native technologies. It has used Red Hat Enterprise Linux for several years, which was one of the driving factors in selecting Red Hat OpenShift over alternate application platform solutions. It has built predictive and prescriptive machine learning models for two primary use cases: loan lending risk management and ATM cash management. Its solutions have been running successfully in production since 2019, and to date, the bank is reporting about 30-40% improvement in overall profits and reduction in risk. By using neural networks, it has been able to identify potential lending risk in its customer base. The use of OpenShift has provided the bank flexibility to extend the solution with its own proprietary or third-party tools (e.g., DevOps for MLOps), as well as supported portability of solutions across different deployment locations. An upgrade to the most recent version of Red Hat OpenShift has provided the bank with better control over security features and adherence to local regulations. Overall, it has helped the bank avoid vendor lock-in and accelerate time to value.

Breaking the cycle of veteran suicide with an open source approach:

The U.S. Department of Veterans Affairs (VA) is looking for innovative ways to address the issue of veteran suicide. In 2022, Red Hat teamed with global consulting services provider Guidehouse and Philip Held, Ph.D. of Rush University Medical Center, to develop new data-driven means of identifying veterans at risk for suicide. The combined group is known as Team Guidehouse. The proposed solution combines the VA's suicide prevention risk model, REACH-VET, with Guidehouse's (in)Sight Health Catalyst, which uses publicly available social media data to identify potentially imminent cases where veterans need immediate suicide intervention support. Leveraging this data, the model seeks to enable providers to identify and help specific veterans in need more easily. The solution depends on the ability to use artificial intelligence and machine learning to sift through vast volumes of data.

Utilizing Red Hat OpenShift Service on AWS (ROSA), Team Guidehouse is able to run the solution natively on Amazon Web Services (AWS), allowing them to increase operational efficiency and benefit from the joint support and operation of Red Hat and AWS. The system relies on Application Programming Interfaces to allow for integration with electronic health records. By leveraging Red Hat OpenShift API Management, Team Guidehouse's solution offers an API-first approach to building microservices applications and helps streamline integration into the existing systems. That ready access to medical histories will be key to help identify those at risk, in support of timely interventions. To help accelerate and amplify machine learning capabilities, the solution also utilizes Red Hat OpenShift AI, an AI platform for data scientists and developers that provides a fully supported environment to rapidly develop, train, and test machine learning models in the public cloud or on-site before deploying into production.

Developing a suicide prevention solution on open source tools not only supports engagement with other partners — extending the solutions as new data sources and insights come to the fore — but also supports a future-ready solution that will be able to continue to adapt and address the challenges in the years to come.

Scaling AI adoption with flexible deployment choices:

A U.S. audit, tax, and advisory firm, one of the world's leading professional services firms, is using a self-managed version of Red Hat OpenShift. It decided to use Red Hat OpenShift because of the beneficial features, namely enterprise-grade scale and security. The firm's AI/ML solutions are not only for its clients in the financial services, healthcare, and insurance industries, but also for its own internal consumption. The solutions have been running successfully in production for over two years and have helped improve operational efficiency while helping end users make time for higher-value work. These solutions empower businesses to extract insights from all types of unstructured content, such as image, voice, or PDF files. The solutions put the documents through a process of digitization, text, and insights extraction using computer vision and natural language processing (NLP) technologies, and therefore superior decision making. They also have conversational AI and machine learning solutions for automating customer services and predictive analytics for the automotive industry. The primary benefits of using Red Hat OpenShift have been scale and portability across cloud providers and in a hybrid cloud setup. OpenShift is also enabling seamless integration into the ecosystem. Overall, Red Hat has been a trusted partner in educating and enabling enterprise requirements.

The Red Hat approach to AI/ML workloads provides flexibility in deployments and helps businesses accelerate time to value with an easy-to-manage and collaborative offering.

Conclusion

AI is becoming ubiquitous across all functional areas of a business. Advancements in ML, NLP, conversational AI, and computer vision AI are at the forefront of AI software innovations. AI initiatives offer more than cost savings: They help organizations predict and shape future outcomes; allow people to do higher-value work; automate and optimize decisions, processes, and experiences; assist in problem resolution, threat mitigation, resource/service optimization, IT staff productivity/efficiency enhancement of high-value tasks, infrastructure dynamics, and digital acceleration; and reimagine new business models. Ultimately, this means increased revenue, improved profit margins, and reduced risks. That said, AI is not a panacea; it holds inherent risks for security, privacy, and equity if not mitigated properly.

As per IDC research, enterprises are looking to accelerate time to value by leveraging AI/ML solutions. Customers need and deserve choice, and open source AI is gaining traction. Building MLOps capabilities on top of an open application platform lets data science professionals and ML practitioners quickly and easily deploy a broad selection of AI/ML tools across hybrid environments in a consistent fashion. Using cloud-native application life-cycle management components such as Kubernetes Operators and Helm Charts, integrated with orchestration tools, lets practitioners directly administer applications, including AI/ML development and deployment tools provided with Red Hat OpenShift AI offerings. IT operations teams can automate common life-cycle management tasks like configuration, provisioning, and updates to gain efficiency, speed, and accuracy. Automation also frees staff from tedious tasks, allowing them to focus on more interesting strategic activities.

The Red Hat approach to AI/ML workloads provides flexibility in deployments and helps businesses accelerate time to value with an easy-to-manage and collaborative offering. Red Hat has an opportunity to simplify and better articulate its commercial terms.

Appendix: Supplemental Data

This appendix provides an accessible version of the data for any complex figures in the document. Click "Return to original figure" to get back to the original data figure.

FIGURE 2 SUPPLEMENTAL DATA

AI Solutions: Leading Use Cases

	Today	Next Two Years
IT automation	13%	12%
Intelligent process automation (business processes)	9%	10%
Profit and loss analysis	9%	8%
Automated threat intelligence and prevention	8%	9%
Automated customer service agents	8%	8%

n = 2,053; Source: IDC's *AI StrategiesView 2022,* CY22

Return to original figure

FIGURE 3 SUPPLEMENTAL DATA

AI/ML Solutions: Deployment Locations

	<999	>1,000
Private cloud	55%	61%
Public cloud	53%	59%
Hybrid cloud	20%	33%
On-premises	28%	32%
Colocation	27%	20%
Edge (on devices like iPhone, iPad, Google Android phone, etc.)	26%	27%
Edge (excludes endpoints)	18%	19%
Multiple	4%	4%

n = 2,053; Source: IDC's *AI StrategiesView 2022,* CY22

Return to original figure

About the IDC Analysts

Ritu Jyoti
Group Vice President, Worldwide Artificial Intelligence and Automation Research Practice, Global AI Research Lead, IDC

Ritu Jyoti is group vice president, covering worldwide artificial intelligence and automation research with IDC's Software Market Research and Advisory practice. Ritu is responsible for leading the development of IDC's thought leadership for AI research and managing the research team. Her research focuses on the state of enterprise AI efforts and global market trends for the rapidly evolving AI and machine learning innovations and ecosystem. She also leads insightful research that addresses the needs of AI technology vendors and provides actionable guidance on how to crisply articulate their value proposition, differentiate, and thrive in the digital era.

More about Ritu Jyoti

Message from the Sponsor

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 integrate new and existing IT applications, develop cloud-native applications, standardize on our industry-leading operating system, and automate, secure, and manage complex environments.

Red Hat OpenShift AI provides tools across the ML lifecycle. Built on top of Red Hat OpenShift, the industry's leading Kubernetes powered hybrid cloud platform, Red Hat OpenShift AI gives data scientists and developers a powerful open hybrid AI/ML platform for gathering insights from data and building ai-enabled applications. In addition, Red Hat actively participates in the Open Data Hub, Kubeflow, and Ray open source communities that provide blueprints, toolkits, and workflows for using Jupyter notebooks, TensorFlow, PyTorch, Kubeflow, and other popular data and ML tools.

To find out more, visit: **red.ht/openshift_ai**

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

≋IDC