

在金融服务中加快 采用人工智能



AI 技术的全球化部署
已达到历史最高水平。

- ▶ 94% 的业务领导者认为 AI 对未来 5 年的成功至关重要。²
- ▶ 79% 的企业部署了 3 种或以上的 AI 应用。²
- ▶ 76% 的领导者计划加大 AI 投入。²

智能金融服务应用为业务成功添砖加瓦

人工智能 (AI) 和机器学习 (ML) 模型使用海量数据来提供有价值的深度见解，自动执行任务并提高核心业务能力。这些技术有可能为金融服务业带来全方位的转变，涵盖从客户和员工到开发和运维等所有领域。当今的金融服务企业使用 AI/ML 模型来开发智能的云原生应用，这些应用可以带来可衡量的结果，如提高客户满意度、提供多样化服务、改进业务和 IT 自动化，以及提升员工效率和生产率等。事实上，36% 的金融服务企业报告称，AI/ML 技术可将年均成本降低 10% 以上。¹

如今，许多金融服务企业在业务单位或部门级解决方案中部署了 AI/ML 模型，以扩充和增强目前由员工执行的活动。例如，ML 算法可以检测人类分析师可能遗漏的异常交易和支出模式，并将其标记以便反欺诈团队开展调查。与此同时，AI/ML 技术还可以分析账户活动，以识别可疑模式，并辨别看似不相关账户之间的联系，这种账户常常参与到复杂洗钱计划中。AI 聊天机器人可以快速、准确地回答常见的支持问题，甚至还能从以前的互动中学习，以改善客户体验，并让员工专注于更复杂的事务。通过将新的 ML 算法应用于经典的光学字符识别 (OCR) 技术，金融机构可以提高了解客户 (KYC) 流程中文件扫描和数字化工作的准确性。

然而，在生产中构建智能应用和部署 ML 运维 (MLOps) 可能困难重重。金融服务企业必须克服重重挑战，才能交付基于 AI/ML 模型的创新服务：

- ▶ AI/ML 专家人数有限，很难找到和留住数据科学家和工程师、ML 工程师、软件开发人员以及其他具备正确知识的工作人员。
- ▶ 数据科学家、ML 工程师、开发人员和其他工作人员之间缺乏联系和协作，减慢了 AI/ML workflows 和模型生命周期。
- ▶ AI/ML 工作流的所有阶段（数据准备和管理、模型训练和微调以及推理等）都需要昂贵的基础架构，这些基础架构可能非常复杂，很难以自动化的自助方式高效地部署和使用。



红帽官方微博



红帽官方微信

¹ NVIDIA, “金融服务行业 AI 现状: 2023 年趋势”, 2023 年。

² Deloitte, “企业中的 AI 现状, 第 5 版”, 2022 年 10 月。



金融服务业热门 AI 用例

金融服务企业成功地将 AI/ML 用于许多用例：¹

- ▶ 自然语言处理和大语言模型
- ▶ 推荐系统和下一代最佳操作系统
- ▶ 投资组合优化
- ▶ 交易和付款欺诈检测
- ▶ 反洗钱和 KYC 计划
- ▶ 算法交易
- ▶ 对话式 AI 系统
- ▶ 营销优化
- ▶ 风险分析

- ▶ AI/ML 模型的训练过程必须遵守数据治理和主权法规，造成数据共享被限制在机构内的某些地理位置或部门。
- ▶ 相较于标准的应用部署，AI/ML 解决方案和管道更为复杂。它们必须支持密集的模式开发、训练和推理过程；与智能应用的简单集成和部署；以及基于观察到的行为来监控、更新和再训练模型的能力。

利用混合云、容器和 Kubernetes 技术克服 AI/ML 挑战

基于 AI/ML 的智能应用需要运用现代的云原生工具和最佳实践来管理开发和部署的复杂性。容器和 Kubernetes 技术提供了部署敏捷性、管理能力和可扩展性，以交付和管理智能应用。作为轻量级的独立软件单元，容器可将应用和依赖项（包括运行时、库、系统工具和设置）打包为易于分发的镜像。Kubernetes 是一个开源容器编排平台，可以在整个环境中自动创建、部署、扩展和管理容器实例。

基于 Kubernetes 的容器编排平台提供了多项优势，为 AI/ML 解决方案奠定了基础。有了自动化和自助服务功能，您可以按需置备 MLOps 环境，以加快和简化 AI/ML 模型的开发、测试和部署。容器是可移植的，因此您可以在不同基础架构占用空间之间一致地使用您的模型，包括物理、虚拟化、私有云、公共云和混合云，无需进行任何更改。Kubernetes 可以自动扩展容器工作负载，以确保应用可用性，同时仅将资源分配给需要它们的工作负载。而且，接着这样一个通用技术平台，您可以毫不费力地将来自开源和商业供应商的强大生态系统中的组件集成到自己的 AI/ML 解决方案中。

虽然如此，但有效的 AI/ML 解决方案需要的不仅仅是标准的 Kubernetes 功能。现代应用平台在 Kubernetes 基础上进行扩充，纳入了高级工具和更强大的功能。诸如 Tekton 和 Jenkins 等持续集成和部署工具可以帮助您快速构建、测试、打包、更新和部署 AI/ML 模型。GitOps 持续交付工具（如 ArgoCD）允许您以代码形式定义复杂的应用部署并实现自动化。标准化的监控服务和应用编程接口（API）将模型中所有部分的信息聚合到一处，让您能够衡量准确性、检测偏差并启动更新或再训练。通过集成通常在虚拟机（VM）中运行的容器与传统应用，您可以将现有工具整合到 AI/ML 解决方案中。此外，与容器存储接口（CSI）支持的广泛数据存储和实时数据馈送建立自助连接，简化并改进了模型训练。

在统一的环境中开发和部署智能应用

红帽® 技术相互结合，构建了一个统一、灵活的 MLOps 基础，为整个 AI/ML 生命周期提供支持，涵盖从模型开发和训练到智能应用集成和部署等各个方面。这一解决方案提供了必要的工具，方便您在所有类型的基础架构（包括物理硬件、虚拟机以及私有云、公共云和混合云）中，以精简、一致的方式构建和部署创新的应用。借助许多经认证的合作伙伴产品和受支持的开源组件，您可以根据企业的需求来自定义环境。



金融服务企业在各种环境中运行 AI/ML 工作负载：

- ▶ 44% 使用混合云环境。¹
- ▶ 32% 仅使用公共云环境。¹
- ▶ 16% 仅使用本地数据中心。¹
- ▶ 5% 使用边缘环境。¹
- ▶ 4% 使用私有云环境。¹



图 1. 红帽将关键技术整合成一个统一的解决方案，以实现一致且标准化的 MLOps 部署。



开源和 AI/ML

世界各地的研究人员和开发人员使用开源开发模型来创建许多最先进的 AI/ML 技术。

作为全球最大的开源公司，红帽相信，开源的开发模式将有助于打造更稳定、更安全、更具创新性的技术。

在红帽，我们协作开展多个社区项目并为开源许可提供保障，以便我们能够继续开发各种软件来解决客户面临的最复杂挑战。

[深入了解](#)我们的开源承诺。

解决方案中的各个层面承担不同的关键职能，以开发、部署和管理 AI/ML 模型和智能应用。

- ▶ **软件基础层**包含操作系统和容器编排工具，为整个 AI/ML 解决方案奠定一致、稳定且注重安全的基础。内置的 **DevOps** 和自动化功能（如自助置备），以及包含的一整套工具、语言和运行时，可以帮助您快速开发和维护创新的、高度分散的云原生 AI/ML 模型和应用。
- ▶ **应用管理层**包含各种工具，可用于在混合云环境中部署和管理云原生应用和服务。监控和策略执行功能可确保应用一致、高效地运行。
- ▶ **数据管理层**包含安全至上并可扩展的数据流工具，以及用于训练和存储 AI/ML 模型的结构化和非结构化数据存储。庞大的可扩展存储使您能够训练最大的 AI/ML 模型并保持高准确率。
- ▶ **内容和编排工具**及技术可帮助您管理 MLOps 环境和工作流。容器镜像仓库以及持续集成和持续部署框架和工作流，可以确保应用构建和部署过程在所有基础架构中保持一致和可重复。
- ▶ **安全和治理层**包括各项技术，可以保护 AI/ML 训练和推理工作负载的安全，防止它们受到未经授权的访问。提高数据中心和区域之间的安全性，有助于保护传输中敏感数据的安全。
- ▶ **人工智能和机器学习技术层**包括用于开发、训练和使用 AI/ML 模型的认证工具和服务。底层平台的自助置备功能可以帮助您以协同、一致的方式从试验快速过渡到生产。

红帽助您打造一个开放、灵活的 MLOps 基础

作为企业开源软件的龙头企业，红帽提供了齐全的产品组合、可靠的专业知识，并与关键的独立软件供应商（ISV）建立了战略合作伙伴关系，帮助您构建 AI/ML 解决方案架构。凭借一系列精挑细选并享有支持的认证开源技术，红帽为构建生产就绪 MLOps 环境提供了一个安全至上的基础。在数据中心和云基础架构之间保持一致，您可以构建 AI/ML 模型和智能应用，提供卓越的洞察力和用户体验。

这一基础的各个组成部分为 MLOps 提供了关键功能。

红帽 OpenShift AI 将红帽 OpenShift 和红帽 OpenShift 数据科学的可靠功能组合成一个一致且可扩展的基础，以实现快速应用创新和跨基础架构协调部署

- ▶ **红帽 OpenShift** 是一个统一的企业就绪型应用平台，专为云原生开发、部署和编排而设计。按需提供计算资源并支持硬件和图形处理单元（GPU）加速，包括 NVIDIA GPU 和英特尔® 深度学习加速（Intel DL Boost），加快了建模和推理任务。在本地、公共云和边缘环境之间保持一致，提供了团队取得成功所需的速度和灵活性。例如，您可以创建一个自助式 MLOps 平台，供数据科学家、数据工程师



客户成功案例亮点：

Banco Galicia

在与红帽咨询团队合作后，Banco Galicia 使用红帽 OpenShift、红帽集成和红帽单点登录技术构建了一个基于 AI 的智能自然语言处理（NLP）解决方案。

关键成果：

- ▶ 验证时间从数日缩短到几分钟，准确率达到 90%
- ▶ 应用停机减少 40%
- ▶ 敏捷性提高了 4 倍

阅读[成功案例](#)。

和开发人员用来快速构建模型，将模型整合到应用中并执行推理任务。借助协作功能，团队可以通过一致的方式创建和共享容器化建模结果。

红帽 OpenShift 包括用于智能应用开发和交付的多个关键工具。[红帽 OpenShift 虚拟化](#) 允许您将现有的虚拟化应用作为原生 Kubernetes 对象进行集成、运行和管理。[红帽 OpenShift Pipelines](#) 允许您在 Kubernetes 原生持续集成和持续部署框架中设计 AI/ML 管道，并在可按需扩展的独立容器中运行每一个管道步骤。另外，[红帽 OpenShift GitOps](#) 提供了一个持续部署工作流，可以根据您在 Git 中创建和存储的配置来自动管理 MLOps 环境。

红帽 OpenShift 随附了[红帽企业 Linux®](#)，可在数据中心、云和边缘环境中提供一致、可扩展和高性能的操作基础。[安全增强型 Linux \(SELinux\)](#) 等内置安全防护功能可帮助您抵御威胁，时刻在行业与监管要求方面保持合规。而且，红帽平台能够彼此协同运行，因此这些功能和认证可以扩展到整个红帽软件堆栈。

- ▶ [红帽 OpenShift 数据科学](#) 是一款基于 [Open Data Hub](#) 开源项目的 AI 平台产品。它为数据科学专家和开发人员提供了一个强大的 AI/ML 平台，以便他们收集深度见解和构建智能应用。该服务包括 Jupyter、TensorFlow 和 Pytorch 等工具和框架，并且集成了 Starburst、Anaconda、IBM 和英特尔等技术合作伙伴的组件，以进一步加快智能金融服务应用的开发。

[红帽集成](#) 是一套全面的集成和消息传递技术，旨在跨混合基础架构连接应用和数据。红帽集成中的关键组件：

- ▶ [红帽应用运行时](#) 是一套包括轻量级运行时和框架的产品、工具和组件，可帮助您开发和维护高度分散的云原生 AI/ML 应用。
- ▶ [红帽 AMQ](#) 是一个灵活并可扩展的分布式消息传递平台，可将数据流式传输到您的 AI/ML 模型，具有高吞吐量和低延迟的特点。

[红帽 OpenShift 数据基础](#) 是一种软件定义存储，可为最大的 AI/ML 数据集提供可扩展的持久型文件、块和对象存储。

[红帽服务互连](#) 是一种 OTT 通信协议，可以简化跨多个数据中心或区域的应用和服务之间的连接。开发团队中的任何人都可以使用该协议，无需提升权限，也不会牺牲安全性。

[红帽 Kubernetes 高级集群管理](#) 是一个统一的控制台，可用来基于云环境中的不同地理位置，大规模地控制、自动化和监控应用部署、集群管理和策略实施。

[红帽 Kubernetes 高级集群安全防护](#) 可以帮助您跨混合云部署保护容器化 Kubernetes 工作负载，以增强应用的安全性。

[红帽 Quay](#) 是一个安全增强的私有镜像仓库平台，可用于跨混合云环境管理云原生内容。

66

“红帽提供的技术和架构，在 Banco Galicia 实施 NLP 的过程中发挥了重要作用。在红帽的支持下，我们成功掌握了红帽 OpenShift 的所有功能。此外，我们还开始着手设计一个可实现云原生的架构。”

Matias Lorusso
Banco Galicia
解决方案架构师

最后，借助我们的**认证合作伙伴生态系统**，您可以将流行的 AI/ML、数据分析、管理、存储、安全和开发工具集成到这个架构中。我们与合作伙伴建立了密切合作，在我们的平台上验证他们的软件，以提高可管理性、安全性和支持性。许多合作伙伴还提供认证的**红帽 OpenShift 操作器**来简化软件生命周期管理。

利用大语言模型

在金融服务行业中，文件数字化、报告分析和会话服务应用通常利用 GPT-4 (Generative Pre-trained Transformer 4) 等大语言模型 (LLM)。由于创建 LLM 需要投入大量工作和计算资源，企业通常在这类应用中使用预训练的模型。即便如此，这些模型仍然需要额外的领域或公司相关的训练，并使用较小的本地数据集进行微调，从而能够提供准确的结果。每个应用都需要一个经过专门训练和调优的专用模型。

红帽 OpenShift AI 是一个用来训练和调优 LLM 的理想平台。借助一个可扩展且一致的平台来进行模型开发、训练和推理以及应用集成和部署，可以消除重复工作，提高整个 AI/ML 生命周期的效率。利用高级集群管理和自助服务功能，您可以构建可在多个模型和应用中重复使用的可重复 AI/ML 管道。事实上，部署红帽 OpenShift AI 的企业在数据科学家工作效率方面提高了 20%。³

了解更多

红帽凭借完整的技术产品组合、可靠的专业知识和战略合作伙伴关系，大力帮助您实现目标。除了用于快速采用的服务和培训外，我们还为构建生产就绪的 AI/ML 环境奠定了基础。无论您处于 AI/ML 旅程的哪一阶段，我们都能帮助您打造一个生产就绪的 MLOps 环境，从而加快智能应用的开发和交付。

- ▶ [进一步了解用于金融服务业的 AI/ML 解决方案。](#)
- ▶ [免费预约业务探讨](#)，详细了解我们可以如何帮助您部署 MLOps 基础。

³ 由 Forrester Consulting 代表红帽进行的委托研究，“[红帽 MLOps 混合云平台的总体经济影响™](#)”，2022 年 3 月。该结果是根据受访客户代表的数据综合整理而成。



关于红帽

红帽是世界领先的企业开源软件解决方案供应商，依托强大的社区支持，为客户提供稳定可靠且高性能的 Linux、混合云、容器和 Kubernetes 技术。红帽致力于帮助客户开发云原生应用，集成现有和新的 IT 应用，并实现复杂环境的自动化和管理。作为深受《财富》500 强公司信赖的技术顾问，红帽旨在提供一流的支持、培训和咨询服务，努力将开放创新的优势赋能于各行各业。红帽作为全球企业、合作伙伴和社区网络的互连枢纽，致力于帮助企业发展、转型，并拥抱数字化未来。



红帽官方微博



红帽官方微信

销售及技术支持

800 810 2100
400 890 2100

红帽北京办公地址

北京市朝阳区东大桥路 9 号侨福芳草地大厦 A 座 8 层 邮编: 100020
8610 6533 9300