



适合所有应用程序的亚马逊云科技

充分发挥生成式人工智能 的潜力：更快、更好地 构建应用程序



目录

创造卓越的成果	3
准备好加入生成式人工智能浪潮了吗?	4
为生成式人工智能选择最具成本效益的云平台	5
将开发人员的工作效率提升到新的水平.....	7
为任何使用案例构建、训练和部署机器学习 (ML) 模型	9
以更强的能力和更低的成本训练和运行模型	11
准备好利用这股不可思议的力量了吗?	12
为什么要使用亚马逊云科技构建?	14

创造卓越的成果

得益于可扩展计算能力的实现、数据的指数级增长以及机器学习（ML）技术的飞速发展，人工智能（AI）的采用正不断增长。

而如今，生成式人工智能正在以更快的速度改变数据格局。

生成式人工智能代表着人工智能领域最前沿的技术，能够在诸多领域创造全新的内容和进行创新。可生成对话、故事、图片、音乐、视频甚至代码，带来无限可能。

根基模型的力量，以及为什么您不需要创建自己的模型

生成式人工智能的核心是根基模型（FM），这是一种具有数十亿个参数的大型模型，已基于大量数据进行了预训练。FM 拥有一系列非凡的功能。此外，这些功能强大的 FM 可以进行量身定制，以用于执行特定于领域的任务，因而，对于试图在竞争激烈的市场中打造差异化竞争优势的企业而言，FM 已成为无价之宝。只需利用一小部分从头开始构建、训练和部署模型所需的数据和计算资源，企业便可定制体现其独特理念、风格和服务的 FM。这种定制化水平使各消费行业的公司能够提供卓越且个性化的客户体验，而无需投入巨资开发自己的 FM。例如，一家金融公司如果想要自动生成日常活动报告，可以基于以往的报告和其他相关的专有数据，对已有的 FM 进行训练，创建一个能够从头开始生成新报告的微调模型（成本远远低于创建新的 FM）。



随着利用自然语言处理技术的工具进入企业和社会，这些工具或将推动未来 10 年内全球 GDP 增长 7%（约合近 7 万亿美元），并推动生产率每年提高 1.5%。

Goldman Sachs²

准备好加入生成式人工智能浪潮了吗？

亚马逊云科技正致力于协助客户快速采用生成式人工智能应用程序，以推动客户实现生产力水平的提升并完成产品/服务转型，同时为下一步计划做好准备。

亚马逊云科技正助力企业克服与这种复杂且要求严苛的技术相关的障碍。亚马逊云科技具备性价比最高的机器学习基础设施，开发人员利用亚马逊云科技能够最便捷地使用 FM 进行构建，能够灵活地使用开源 FM 或构建自己的模型，还能够使用 Amazon CodeWhisperer 等由 FM 提供支持的工具来快速提高工作效率。

通过采用亚马逊云科技基于云的解决方案，开发人员可以利用创新的生成式人工智能产品来满足客户日益增长的需求。在本文档中，我们将深入探讨这些专用工具和服务，探索其独特功能，并展示其如何在全新的生成式人工智能领域助力您提升业务水平。

McKinsey 的一份报告指出，90% 的商业领导者希望在未来两年内定期使用生成式人工智能产品。¹

为生成式人工智能选择最具成本效益的云平台

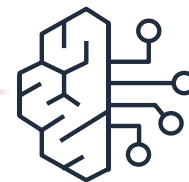
在超过 25 年的时间里，亚马逊云科技一直在利用人工智能提供卓越的客户体验，从 [Amazon Prime Air](#) 无人机送货系统到 [Amazon Go](#) 商店和 [Alexa](#)，不一而足。亚马逊云科技已协助超过 10 万家各种规模和行业的客户利用这一变革性技术，其中包括 Intuit、Thomson Reuters、AstraZeneca、Ferrari、Bundesliga、3M 和 BMW，以及全球数以千计的初创企业、中小型企业（SME）和政府机构。

亚马逊云科技对生成式人工智能技术采用同样的普及方式：试图将这种技术带出研发领域，让更多企业和大众可以应用此项技术，而不是只有少数初创企业和资金雄厚的大型科技公司才能触及。借助基于亚马逊云科技的生成式人工智能，您能够革新应用程序，打造全新的客户体验，推动生产力达到前所未有的水平，并实现业务转型。

亚马逊云科技提供了一套不断改进的产品和服务，使企业能够充分利用生成式人工智能的强大功能。您可以从一系列常用的 FM 中进行选择，也可以使用内置生成式人工智能的亚马逊云科技服务，所有这些服务都在最具成本效益的生成式人工智能云基础设施上运行。

使用根基模型构建应用程序的最便捷场所

[Amazon Bedrock](#) 是客户使用 FM 构建和扩展生成式人工智能应用程序的最简便方法，并在所有构建者之间推广普及。Amazon Bedrock 让用户可通过 API 访问亚马逊和领先的人工智能初创企业（包括 AI21 Labs、Anthropic、Cohere 和 Stability AI）提供的 FM。



普及高性能 FM 的使用



客户往往寻求采用高性能的 FM，以期根据自身的具体需求取得卓越的业务成果。Amazon Bedrock 使管理大型基础设施集群不再复杂，并大幅降低了成本。通过与应用程序无缝集成，Amazon Bedrock 使用户能够毫不费力地构建和扩展生成式人工智能应用程序，同时又不会影响性能。

利用专有数据安全构建差异化应用程序



借助 Amazon Bedrock，客户可以使用自己的数据轻松构建定制的应用程序，同时确保数据的安全性和隐私性。Amazon Bedrock 让用户能够安全地基于基础 FM 进行构建，确保用户的应用程序在市场中脱颖而出。通过安全地利用专有数据，用户能够创建差异化应用程序来满足自己的独特要求。

轻松集成和部署



Amazon Bedrock 通过无服务器体验简化了模型选择过程，提供了可扩展性和灵活性，同时消除了基础设施管理任务。用户能够轻松找到适合自己需求的理想模型，并快速上手。借助 Amazon Bedrock，客户能够使用自己的数据专门定制自己的 FM，并能够使用熟悉的亚马逊云科技工具和功能，将 FM 无缝集成和部署到应用程序中。用户能够利用 [Amazon SageMaker](#) 的机器学习功能，如用于测试不同模型的 Experiments 和用于大规模管理 FM 的 Pipelines，所有这些都不需要管理任何基础设施。

最前沿的 FM 触手可及



Amazon Bedrock 提供了当今最前沿的一些 FM。例如，客户能够利用 AI21 Labs 的 [Jurassic-2](#) 系列的多语言大型语言模型 (LLM)，并能够根据自然语言指令生成西班牙语、法语、德语、葡萄牙语、意大利语和荷兰语的文本。此外，Anthropic 最新的大型语言模型 [Claude 2](#) 基于 Constitutional AI 和无害化训练，在思维对话、内容创作、复杂推理、创造力和代码编写方面表现超群。Cohere 的 [Command](#) 文本生成模型是面向实际业务应用而训练的，而 Embed 模型则是面向 100 多种语言的搜索、聚类或分类任务而训练的。Stability AI 的文本到图像根基模型（包括广受欢迎的 [Stable Diffusion](#)）可生成独特、逼真、优质的图像、艺术作品、徽标和设计作品。

定制变得简单



Amazon Bedrock 最显著的特点之一就是易于定制。客户的任何数据都不会用于训练原始基础模型。客户能够配置自己的 [Amazon Virtual Private Cloud \(Amazon VPC\)](#) 设置来访问 Amazon Bedrock API，并以安全的方式提供微调的模型数据，而且所有数据都经过加密。无论是传输中 (TLS1.2) 的客户数据还是静态数据，会始终通过服务托管密钥进行加密。

早期用户对 Amazon Bedrock 大加赞赏

Amazon Bedrock 已经引起了 Coda 等客户的极大兴趣。Coda 联合创始人兼首席执行官 Shishir Mehrotra 在评价 Amazon Bedrock 时，称赞它使得公司的人工智能解决方案实现了优异的质量、可扩展性和性能，同时保证了数据的安全性。Amazon Bedrock 的便利性激发了各种规模的企业对它的热情，包括 Accenture、Deloitte、Infosys 和 Slalom 等知名合作伙伴，这些企业正在摸索实操方法来加快生成式人工智能的采用。C3 AI 和 Pega 等独立软件供应商 (ISV) 也在利用 Amazon Bedrock，对于 Amazon Bedrock 丰富的 FM 选择以及亚马逊云科技在安全性、隐私性和可靠性方面的承诺，这些企业深表赞赏。

将开发人员的工作效率提升到新的水平



亚马逊云科技能够通过多种方式让您充分体验生成式人工智能的进步。其中一种方法是，将生成式人工智能融入到有助于加速和简化日常任务的服务中。例如，Amazon CodeWhisperer 是一款颠覆性的人工智能代码编写工具，正在彻底改变开发人员的工作效率。



利用生成式人工智能变革开发人员的体验

在使用 [Amazon CodeWhisperer](#) 之前，软件开发人员需要花费大量时间编写代码。而有了 CodeWhisperer，开发人员现在能够在集成式开发环境（IDE）中利用人工智能生成的代码建议。通过将更多的代码编写工作交给 CodeWhisperer，开发人员能够加快开发速度，释放创造力。客户对 CodeWhisperer 的积极反馈证明该工具有潜力成为最强大的生成式人工智能应用程序之一。



内置了安全扫描功能的人工智能代码编写工具

如果生成式人工智能工具建议的代码包含隐藏的安全漏洞，或者不能负责任地处理开源代码，那么开发人员的工作效率就无法真正得到提高。因此 CodeWhisperer 还内置了安全扫描功能，由自动推理功能提供支持。该工具能识别难以检测到的漏洞并提出修复建议，包括那些被列入开放式全球应用程序安全项目（OWASP）十大威胁中的漏洞，以及不符合加密库最佳实践的漏洞等。CodeWhisperer 还能过滤掉有偏见或不公平的代码建议，并能标记出类似于开源代码的代码，供客户参考或授权使用。



基于亚马逊云科技的最佳代码编写助手

除了从数十亿行公开可用的代码中学习外，CodeWhisperer 还针对亚马逊代码进行了训练。我们相信，CodeWhisperer 现在是为 [Amazon Elastic Compute Cloud \(Amazon EC2\)](#)、[Amazon Lambda](#) 和 [Amazon Simple Storage Service \(Amazon S3\)](#) 等亚马逊云科技服务生成代码最准确、最快速且最安全的方法。

通过 Amazon Toolkit IDE 扩展，CodeWhisperer 可与 VS Code、IntelliJ IDEA、Amazon Cloud9 等受欢迎的 IDE 无缝集成。

CodeWhisperer 支持 15 种语言，包括 Python、Java、JavaScript 等。



面向个人开发人员的免费套餐

为了确保尽可能多的开发人员都能使用 CodeWhisperer，亚马逊云科技为个人用户提供了免费套餐。生成代码没有资格或时间限制；任何人只需一个电子邮件账户即可注册。不需要亚马逊云科技账户。对于企业用户，我们提供了 CodeWhisperer 专业版套餐，其中包括多种管理功能，例如与 [Amazon Identity and Access Management \(IAM\)](#) 集成的单点登录（SSO）等，还有更严格的安全扫描限制。

CodeWhisperer 将 Accenture 开发人员的工作效率提高了 30%

“Accenture 正在使用 Amazon CodeWhisperer 加快代码编写速度，践行我们 Velocity 平台的软件工程最佳实践计划 [...] 在探索了多种选择后，我们遇到了 Amazon CodeWhisperer，**这款工具使我们的开发工作减少多达 30%**，现在，我们将更多精力集中在提高安全性、质量和性能方面。”

Accenture 技术架构高级经理
Balakrishnan Viswanathan

[阅读完整案例 >](#)

为任何使用案例构建、训练和部署机器学习（ML）模型



Amazon SageMaker 建立在亚马逊云科技 20 年来开发实际机器学习应用程序的经验之上，其中涵盖产品推荐、个性化、智能购物、机器人和语音辅助设备。这项独特的服务提供了一套广泛的机器学习功能，成千上万的客户使用这些功能来访问和分析数据，并构建、训练和部署高质量的机器学习模型。

借助 SageMaker，您可以：

- 让更多人能够通过精选工具，利用机器学习进行创新。这些工具包括面向数据科学家的集成式开发环境（IDE）和面向业务分析师的无代码界面。
- 访问、标注和处理大量结构化数据（表格数据）和非结构化数据（照片、视频、地理空间信息和音频）以进行机器学习。
- 利用优化版基础设施，将训练时间从数小时缩短至数分钟。通过专用工具，将团队生产力提升多达 10 倍。
- 在整个企业内实现机器学习运维实践和治理的自动化和标准化，从而提高透明度和可审核性。
- [Amazon SageMaker JumpStart](#) 是一个机器学习中心，可在此处访问算法、模型和机器学习解决方案，从而让您能够快速开始使用机器学习。有了 SageMaker JumpStart，机器学习从业者可以从广泛的[公开可用的根基模型](#)中进行选择。机器学习从业者可以通过网络隔离环境将 FM 部署到专用的 SageMaker 实例，并使用 SageMaker 定制模型，以训练和部署模型。

10 倍

团队生产力提高幅度

1 万亿次
以上

预测数量/月

54%

TCO 降低幅度

高达 50%

训练速度提高幅度（通过更有效地利用 GPU 实现）

40%

标注成本降低幅度

22

项合规性计划（PCI、HIPAA、SOC 1/2/3、FedRAMP、ISO 等）

<10ms

推理开销延迟

面向业务分析师的 SageMaker

- 使用点击式界面轻松准备数据、训练模型并生成预测结果
- 与数据科学家共享模型和数据集，加强合作
- AutoML 集成到常见的 BI 工具中，如 Domo、Snowflake 和 [Amazon Redshift](#)

面向数据科学家的 SageMaker

- 访问来自结构化和非结构化数据来源的数据
- 利用专用工具提高工作效率
- 只需点击几下即可使用完全托管式 Jupyter Notebook

面向机器学习运维的 SageMaker

- 创建可重复的训练工作流程，以加快模型开发速度
- 为机器学习构件集中编目，以实现模型的可重复性和管理
- 将机器学习工作流与 CI/CD 管道集成，以加快生产速度
- 持续监控生产中的数据和模型，以保持质量

观看：[使用 Amazon SageMaker 构建生成式人工智能应用程序 – 亚马逊云科技虚拟研讨会](#)

技术创新研究所在 SageMaker 上训练包含 400 亿参数的 LLM

阿拉伯联合酋长国的技术创新研究所 (TII) 是研究变革性技术的领先研究中心。利用 Amazon SageMaker, TII 能够为开源的 Falcon 模型预处理和训练数据。SageMaker JumpStart 提供 Falcon-40B 和 Falcon-7B 这两种模型。

[阅读完整案例](#) ›

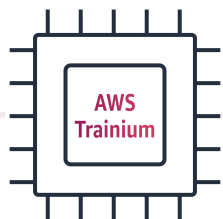
Forethought 利用 SageMaker 将人工智能模型的成本降低 66%

为了让支持团队利用生成式人工智能软件削减成本, Forethought 转而使用亚马逊云科技和 Amazon SageMaker, 来更有效地利用 GPU 等硬件资源。有了 SageMaker, Forethought 能够使用多模型终端节点 (MME), 在单个推理节点上运行多个人工智能模型并扩大规模, 进而避免这种规模的训练往往会产生的高昂成本。

[阅读完整案例](#) ›

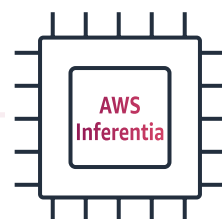
以更强的能力和更低的成本训练和运行模型

生成式人工智能需要强大的处理能力。亚马逊云科技在开发自己的硅芯片方面进行了大量投资，以便在要求苛刻的工作负载中挑战更高的性能和性价比。Amazon Trainium 和 Amazon Inferentia 芯片针对在云中训练模型和运行推理提供了最具成本效益的解决方案。



利用 Trn1 实例实现无与伦比的训练性能

由 [Amazon Trainium](#) 提供支持的 [Amazon EC2 Trn1](#) 实例在机器学习训练中引入了关键的创新。与其他 EC2 实例相比，这些实例可节省多达 50% 的成本，具有卓越的价值主张。Trn1 实例通过第二代 [Elastic Fabric Adapter \(EFA\)](#) 网络，以 800 Gbps 的速度连接多台服务器，对分布式训练进行优化，从而加快训练 workflow。Helixon、Money Forward 和 Amazon Search 团队等客户极大地缩短了大规模深度学习模型的训练时间，将训练时间从几个月缩短到几周甚至几天。有了此番壮举的铺垫，亚马逊云科技正式推出了 Trn1n 实例，这种实例采用 1600 Gbps 的网络带宽，使网络密集型模型的性能提高了 20%。



利用 Inf2 实例为大规模生成式人工智能应用程序提供支持

[Amazon EC2 Inf2](#) 实例由 [Amazon Inferentia2](#) 提供支持，专为大规模生成式人工智能应用程序而设计，所采用的模型具有数千亿个参数。与上一代实例相比，这种实例的吞吐量可提高多达 4 倍，延迟时间可缩短多达 10 倍。通过加速器之间的超高速连接，Inf2 实例可促进大规模的分布式推理。与其他 EC2 实例相比，这些先进技术使推理性价比提升高达 40%，从而最大限度降低了云推理成本。像 Runway 这样的客户在使用 Inf2 实例后，吞吐量提高多达 2 倍，让他们得以推出更多功能和复杂的模型，同时为数以百万计的创作者提供卓越的体验。

准备好利用这股不可思议的力量了吗？

探索生成式人工智能领域的使用案例

有了亚马逊云科技工具和服务，生成式人工智能可发挥巨大的潜力。以下列举了一些行业及其关键使用案例：



金融服务

欺诈检测

识别异常情况、异常交易模式和潜在风险，助力金融机构打击欺诈行为。

产品组合优化

分析大量财务数据，协助资产管理者根据风险状况、收益预期和市场形势做出明智决策。



制造

产品设计优化

生成优化的产品设计，提高制成品的效率和性能。

流程优化

分析生产数据，找出瓶颈、低效原因，并把握机遇，以便最大限度提高生产力。



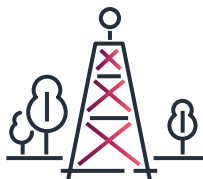
健康与生命科学

药物研发

生成新的分子结构，预测其性质，优化候选药物，并加速药物开发过程。

医学影像分析

分析 X 射线、CT 扫描或核磁共振扫描等图像，以检测异常、肿瘤或其他病症，并改进诊断工作。



电信

营销偏好

即时回复客户询问，既准确又高度个性化。

网络性能

识别潜在的网络问题，推荐故障排除措施，并自动执行优化和维护任务。



政府

公共安全与保障

分析监控录像或传感器数据，以检测异常活动、识别潜在威胁或监控公共安全。

政策模拟与规划

模拟各种情景，预测政策决策的结果，助力政府做出明智的决策。



一般性问题

聊天机器人和虚拟助理

提供类似于真人坐席的自动化客户支持，回答常见问题并处理基本任务。

语言翻译与摘要

在不同语言之间进行翻译或者总结冗长的文档，以实现全球范围的交流、内容本地化和信息提取。

为什么要使用亚马逊云科技构建？

我们具有成套的专用生成式人工智能工具和服务，能够提供最全面、性能最高且高度可扩展的基础设施，让企业实现极具成本效益的机器学习训练和推理。客户如果选择亚马逊云科技来利用生成式人工智能，将会从以下方面受益：



使用 FM 进行构建的最便捷场所

利用熟悉的控件，以及与广泛且深入的亚马逊云科技功能和服务（如 SageMaker 和 Amazon S3）的集成，将 FM 轻松地集成和部署到您在亚马逊云科技上运行的应用程序和工作负载中。



性价比最高的机器学习基础设施

利用由亚马逊云科技设计的机器学习芯片和 NVIDIA GPU 提供支持的基础设施，可获得最具性价比的生成式人工智能。经济高效地扩展基础设施，训练和运行包含数以千亿计个参数的 FM。



灵活地从头开始构建

从 AI21 Labs、Anthropic、Cohere 和 Stability AI 提供的各种 FM 中进行选择，为您的使用案例找到理想的模型。



生成式人工智能驱动的方案

借助内置的生成式人工智能，CodeWhisperer（人工智能代码编写助手）等服务有助于您提高工作效率。此外，您可以使用将亚马逊云科技人工智能服务与领先的 FM 相结合的亚马逊云科技示例解决方案，部署常见的生成式人工智能使用案例，例如通话摘要和问答等。



安全定制

只需几个带标注的示例，即可为您的企业定制 FM。由于所有数据均经过加密，且不会离开您的 Amazon VPC，因此您尽可放心，我们将会始终保护您数据的私密性和机密性。

亚马逊云科技让明天更美好

亚马逊云科技提供了一套集成的服务，让企业可以利用 LLM 和 FM 构建和部署生成式人工智能应用程序。Amazon SageMaker 有助于开发人员训练、调整和优化 LLM，而 Amazon Bedrock 则提供一个控制面板来管理和监控模型。使用 Amazon CodeWhisperer，代码编写人员可以绕过耗时的代码编写任务，并使用不熟悉的 API 加速构建。

凭借端到端功能，亚马逊云科技可助力任何企业快速开发准确、功能强大的生成式人工智能应用程序。

[详细了解基于亚马逊云科技的生成式人工智能](#) ›

¹ 《AI-powered marketing and sales reach new heights with generative AI》，McKinsey，2023 年 5 月

² 《Generative AI could raise global GDP by 7%》，Goldman Sachs，2023 年 4 月