亚马逊云科技

生成式人工智能赋能所有云原生企业

利用亚马逊云科技轻松构建和扩展生成式人工智能

目录

引言

为云原生企业解锁生成式人工智能的力量和前景

机器学习（ML）范式转变的种子早在几十年前就已萌芽，但随着可扩展计算容量的推出、数据的激增以及机器学习技术的快速发展，各行各业的客户都在进行业务转型。OpenAI 的 ChatGPT 和 Google 的 Bard 等生成式人工智能工具受到广泛关注，投资需求不断增长。根据 2023 年 3 月的一份 PitchBook 报告，风险投资家们（VC）正在增加他们在生成式人工智能领域的投资，从 2018 年的 4.08 亿美元增加到 2021 年的 48 亿美元和 2022 年的 45 亿美元。天使和种子交易也有所增长，2022 年投资达到 3.583 亿美元 ，而 2018 年仅为 1.028 亿美元。[1]

考虑到生成式人工智能的益处已有实证，这些风险投资数字并不令人惊讶。作为企业家，您可以借助该技术来自动执行任务、提供个性化的客户体验并优化成本。超过 60% 的企业主已经相信人工智能（AI）将会提高他们的生产力，现在是时候加入他们的行列了。[2]

对于有兴趣将生成式人工智能解决方案集成到各自业务中的云原生企业领导者，这本电子书可作为他们的指南。电子书中包括利用生成式人工智能的云原生企业示例，并说明了为什么各种规模的组织都选择亚马逊云科技来完成生成式人工智能之旅。首先，我们来了解一下该技术的基本原理。

[1] “Vertical Snapshot: Generative AI”（垂直领域速览：生成式人工智能），PitchBook，2023 年 3 月
[2] Haan, K.，“24 Top AI Statistics and Trends In 2023”（2023 年 24 大人工智能统计数据和趋势），福布斯，2023 年 4 月

了解生成式人工智能

生成式人工智能是一种可以生成新内容和想法（包括对话、故事、图像、视频和音乐）的人工智能，由基于大量数据进行预训练的大型模型提供支持，这些模型通常称为根基模型（FM）。

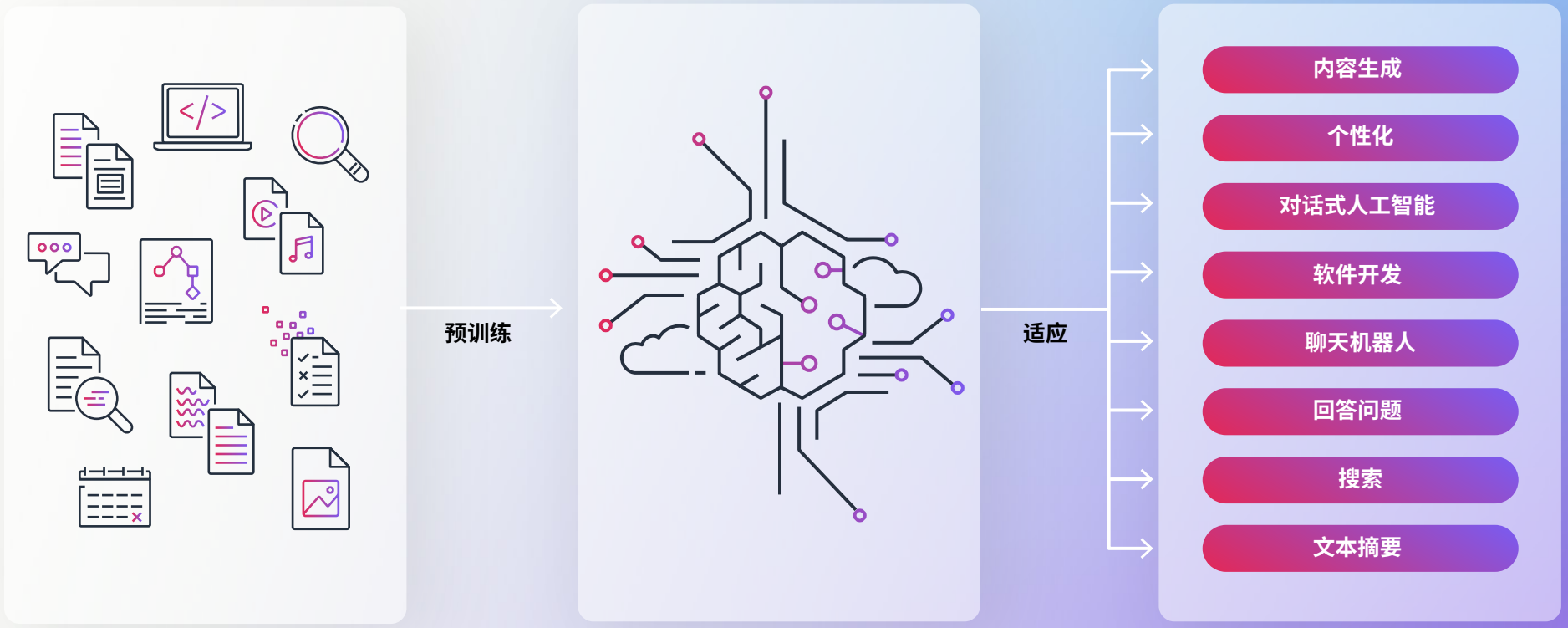
机器学习的最新进展，特别是基于转换器的神经网络架构的发明，带动了包含数十亿个参数或变量的模型的兴起。举例来说，2019 年最大的预训练模型有 3.3 亿个参数。到 2023 年，最大的模型包含超过 5000 亿个参数，在短短几年内增加了 1600 倍。当今的 FM，例如大型语言模型（LLM）GPT3.5 或 BLOOM 以及文本到图像模型 Stable Diffusion，可以执行跨越多个领域的各类任务，例如撰写博客文章、生成图像、解决数学问题、参与对话，以及根据文档回答问题。FM 的大小和通用性质使其不同于传统的机器学习模型，后者通常执行特定任务，例如分析文本中所表达的情绪、对图像进行分类和预测趋势。

通过以各种形式和模式预训练互联网规模的数据，FM 学会了在各种环境中应用所学到的知识。预训练 FM 的功能令人惊叹，同样令人兴奋的是，这些模型还可以定制，用于执行特定领域的功能，而这些功能仅使用从头开始训练模型所需的一小部分数据和计算。

利用定制 FM 提升客户体验

定制 FM 可以打造独特的客户体验，体现公司在各行各业的心声、风格和服务。例如，一家需要使用所有相关交易自动生成日常活动报告的金融科技云原生企业，可以使用专有数据自定义模型。此数据包括过去的报告，使 FM 能够了解应该如何阅读报告，以及生成报告时使用了哪些数据。

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，推动生产力水平达到前所未有的水平，以及实现云原生企业转型。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。



将生成式人工智能应用于您的云原生企业的主要方法

内容生成

协助完成一些任务，例如撰写论文、报告、电子邮件、概念艺术和设计或生成仅靠人力可能无法完成的独特内容。

个性化

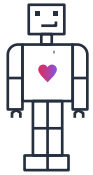
通过为您的网站和通信提供高度相关的内容和产品推荐，为您的客户提供更加个性化的体验。

对话式人工智能

创建基于自然语言的对话界面，例如聊天机器人和虚拟助理，并利用语音转文本和翻译功能。

软件开发

基于自然语言输入生成代码片段、注释和文档，以提高软件开发任务的效率和准确性。



聊天机器人

创建基于自然语言的对话界面，通过提供更人性化的交互来增强用户体验。

回答问题

使用来自大量数据（例如互联网）的自然语言提示查找和整合信息，以及快速回答客户问题。

搜索

在文档和其他资产中查找内容和信息，从而提高搜索准确性、更快地生成搜索结果，以及获取洞察，来制定以数据为依据的业务决策。

文本摘要

生成文章、文档或网页的较短版本；大量文本的简要概述；或一段文字的关键点。

为何选择亚马逊云科技实现生成式人工智能?

20 多年来，人工智能和机器学习一直是亚马逊关注的焦点，客户在亚马逊上使用的很多功能都是由机器学习驱动的。我们的电子商务推荐引擎、履单中心对机器人拣选路线的优化路径，以及我们的供应链、预测和容量规划都由机器学习提供信息和驱动。

Prime Air（我们的无人机）和 Amazon Go（我们的实体零售体验，让消费者可以从货架上挑选商品，然后离开商店，而无需正式结账）中的计算机视觉（CV）技术均采用深度学习（DL）。Alexa 由 30 多个不同的机器学习系统提供支持，每周协助客户数十亿次来管理智能家居、购物、获取信息和娱乐等。我们在亚马逊有成千上万的工程师致力于机器学习，这是我们过往传统、当前宗旨和未来工作的重要组成部分。

亚马逊一直专注于人工智能和机器学习 **20** **余年**

在各种规模的云原生企业中大规模普及生成式人工智能

我们的方法是大规模普及生成式人工智能：我们致力于将这些技术从研究和实验领域中解放出来，将它们推广到少数云原生企业和资金雄厚的大型科技公司之外。客户将亚马逊云科技用于生成式人工智能应用程序有几个关键原因。

1. **最具成本效益的基础设施**：要通过生成式人工智能实现您的目标，您需要专为机器学习构建的性能最高、最具成本效益的基础设施。在过去五年中，亚马逊云科技一直在投资我们自己的硅芯片，力求在机器学习训练和推理等要求苛刻的工作负载的性能和性价比方面实现突破。我们的 **Amazon Trainium** 和 **Amazon Inferentia** 芯片为在云中训练模型和运行推理提供了最低的成本。借助由 NVIDIA GPU 和亚马逊云科技设计的机器学习芯片提供支持的机器学习基础设施，客户可以灵活地选择最佳基础设施，从而在控制成本的同时，更大限度地提高性能。
2. **灵活性**：从领先的 AI 云原生企业和亚马逊提供的广泛模型中进行选择，来满足您独特的业务需求，以及从 AI21 Labs、Anthropic、Stability AI 和亚马逊的多种 FM 中进行选择，从而找到适合您的用例的模型。任何其他供应商都无法提供如此广泛而深入的选择。
3. **安全定制**：只需几个带标签的示例即可为您的企业定制 FM。由于所有数据均被加密且不会离开您的 **Amazon Virtual Private Cloud**（Amazon VPC），您可以相信您的数据能够保持私密性和机密性。亚马逊云科技提供 300 种安全、合规性和治理服务和功能，为客户提供构建适合他们的端到端安全策略所需的灵活性和定制化。
4. **使用 FM 进行构建的最快方式**：将 FM 快速集成并部署到在亚马逊云科技上运行的应用程序和工作负载中。使用熟悉的控件，以及与 **Amazon SageMaker** 和 **Amazon Simple Storage Service**（Amazon S3）等功能和服务的广度和深度的集成。
5. **生成式人工智能驱动的解决方案**：借助内置的生成式人工智能，人工智能代码编写助手 **Amazon CodeWhisperer** 等服务有助于您提高生产力。此外，亚马逊云科技示例解决方案结合了亚马逊云科技人工智能服务与领先的 FM 模型，您可以使用此类方案来部署常见的生成式人工智能用例，例如呼叫汇总和回答问题。

亚马逊云科技实现负责任的人工智能的方法

亚马逊云科技在其全面开发流程的每个阶段都会谨记负责任的人工智能，以此构建 FM。在整个设计、开发、部署和运营过程中，我们考虑了一系列因素，包括：

准确度

评估事实的正确性或摘要反映基础文档的接近程度

公平性

系统如何影响不同的用户亚群（例如，按性别、种族）

知识产权（IP）和版权注意事项

适当用途

过滤掉用户对法律建议、医疗诊断或非法活动的请求

毒性

限制仇恨言论、亵渎、暴力以及攻击性和不当语言

隐私

保护个人信息和客户提示

为解决这些问题，我们将解决方案构建到我们获取训练数据的过程、FM 本身，以及我们用来预处理用户提示和后处理输出的技术中。我们大力投资，致力于改进我们所有 FM 的功能，并在客户尝试新用例时向客户学习。在亚马逊云科技，我们知道生成式人工智能技术及其用途将继续演变，带来需要额外关注和化解的新挑战。我们与学术界、工业界和政府合作伙伴一起，致力于以负责任的方式持续开发生成式人工智能。

详细了解亚马逊云科技实现负责任的人工智能和机器学习的方法 ›

如需详细了解这些挑战和新兴解决方案，请阅读这篇亚马逊科学博客 ›

用于基于亚马逊云科技构建生成式人工智能的工具

1. Amazon Bedrock

使用 FM 构建和扩展生成式人工智能应用程序的最轻松方式

Amazon Bedrock 是一项完全托管式服务，可通过 API 提供来自领先 AI 云原生企业和亚马逊的 FM，因此您可以从各种 FM 中进行选择，从而找到最适合您的云原生企业用例的模型。借助 Bedrock 无服务器体验，您可以快速上手，使用您自己的数据私下定制 FM，并使用您熟悉的亚马逊云科技工具和功能轻松地将 FM集成和部署到您的应用程序中。这些功能包括与 **SageMaker** 的集成、用于测试不同模型的 SageMaker Experiments，以及用于大规模管理 FM 而无需管理任何基础设施的 SageMaker Pipelines 等等。

借助 Bedrock，您可以构建和扩展生成式人工智能应用程序，这些应用程序可以根据提示生成文本和图像。您的团队将获得对重要 AI 云原生企业（包括 AI21、Anthropic 和 Stability AI）的 FM 的访问权，以及对亚马逊云科技开发的 **Amazon Titan** 系列 FM 的专享访问权。

详细了解 Amazon Bedrock ›

stability.ai

Stability AI 是热门图像生成模型 FM Stable Diffusion 背后的开源生成式人工智能公司。“通过 Amazon Bedrock 向亚马逊云科技客户提供我们的 Stable 开放模型套件，从而巩固我们与亚马逊云科技的持续合作关系，这令我感到很欣慰。与亚马逊云科技的这次合作证明了我们致力于提供尖端的开放式人工智能解决方案，从而使企业能够制定更明智的决策，并在不断变化的世界中实现更出色的稳定性。我们相信，这种伙伴关系将为亚马逊云科技客户带来巨大价值，我们期待彼此密切合作，让更广泛的受众能够使用这些强大的功能。”

Emad Mostaque，Stability AI 首席执行官

2. Amazon Titan 模型

借助亚马逊的高质量 FM 进行负责任的创新

Amazon Titan 目前由两个 FM 组成。第一个是 Titan Text，这是一个生成式 LLM，用于诸如汇总、文本生成（例如创建博客文章）、分类、开放式问答和信息提取等任务。第二个是 Titan Embeddings，这个 LLM 会将文本输入（单词、短语或大文本单元）转换为包含文本语义的数字表示（称为嵌入）。虽然此 LLM 不会生成文本，但其对个性化和搜索等应用程序很有用，因为通过比较嵌入，该模型将产生比单词匹配更相关和上下文相关的响应。事实上，Amazon.com 的产品搜索功能使用类似的嵌入模型，来协助客户找到他们正在寻找的产品。为了继续支持负责任地使用 AI 的最佳实践，Amazon Titan FM 旨在检测和删除数据中的有害内容，拒绝用户输入中的不当内容，并过滤包含不当内容的输出，例如仇恨言论、亵渎和暴力。

详细了解 Amazon Titan 模型 ›

3. Trainium 和 Inferentia

高性能、经济高效的生成式人工智能基础设施

Amazon Elastic Compute Cloud（Amazon EC2）Trn1 实例由 **Trainium** 加速器提供支持，专为生成式人工智能模型（包括 LLM 和潜在扩散模型）的高性能深度学习训练而构建。与其他类似的 Amazon EC2 实例相比，Trn1 实例可节省高达 50% 的训练成本。您可以使用 Trn1 实例，在文本摘要、代码生成、回答问题、图像和视频生成、推荐和欺诈检测等广泛的应用情形中训练深度学习模型。

Amazon Neuron SDK 可协助开发人员在 Amazon Trainium 上训练模型，并在 **Inferentia** 加速器上部署模型。它与 PyTorch 和 TensorFlow 等框架原生集成，因此，您可以继续使用现有的代码和工作流在 Trn1 实例上训练模型。

详细了解 Amazon EC2 Trn1 实例 ›

Money Forward

“我们在 Amazon EC2 Inf1 实例上推出了大规模人工智能聊天机器人服务，与基于 GPU 的同类实例相比，推理延迟缩短了 97%，并降低了成本。由于我们需要定期微调定制的 NLP 模型，因此减少模型训练时间和成本也很重要。根据我们在 Inf1 实例上成功迁移推理工作负载的经验，以及在基于 Amazon Trainium 的 EC2 Trn1 实例上取得的初步工作成果，我们预计，Trn1 实例将会在提高端到端机器学习性能和降低成本上给我们带来更多的价值。”

Takuya Nakade，Money Forward, Inc. 首席技术官

Amazon EC2 Inf2 实例由 Amazon Inferentia2 提供支持

Amazon EC2 Inf2 实例专为深度学习推理而构建。此类实例在 Amazon EC2 中以最低成本为生成式人工智能模型（包括 LLM 和视觉转换器）提供高性能推理。您可以使用 Inf2 实例运行推理应用程序，用于文本摘要、代码生成、视频和图像生成、语音识别、个性化、欺诈检测等用途。

Inf2 实例由第二代 Inferentia 加速器 Amazon Inferentia2 提供支持，可提高 Inf1 的性能，将吞吐量提高多达四倍，将延迟缩短多达至十分之一。Inf2 实例是 Amazon EC2 中的首个推理优化实例，可通过加速器之间的超高速连接支持横向扩展分布式推理。您的团队现在可以高效且经济地在 Inf2 实例上跨多个加速器部署具有数千亿个参数的模型。Inf2 实例的性价比比同类 Amazon EC2 实例高出 40%。

详细了解 Amazon EC2 Inf2 ›

Finch Computing 因将 Amazon Inferentia 用于语言翻译而得以节省 80% 的推理成本

Finch Computing 开发自然语言处理（NLP）技术，从大量文本数据中挖掘洞察信息，希望满足客户支持更多语言的请求。Finch 使用深度学习算法构建了自己的神经翻译模型，计算要求很高，并依赖于 GPU。现在，该公司需要一个可扩展的解决方案来支持全球数据源，并使其能够快速迭代新的语言模型，而不会产生高昂的成本。该公司创建了一个以 Amazon Inferentia 的使用为中心的计算基础设施。基于这一点，Finch 缩短了产品上市时间，将其 NLP 扩展为支持三种新语言，并吸引了新客户。

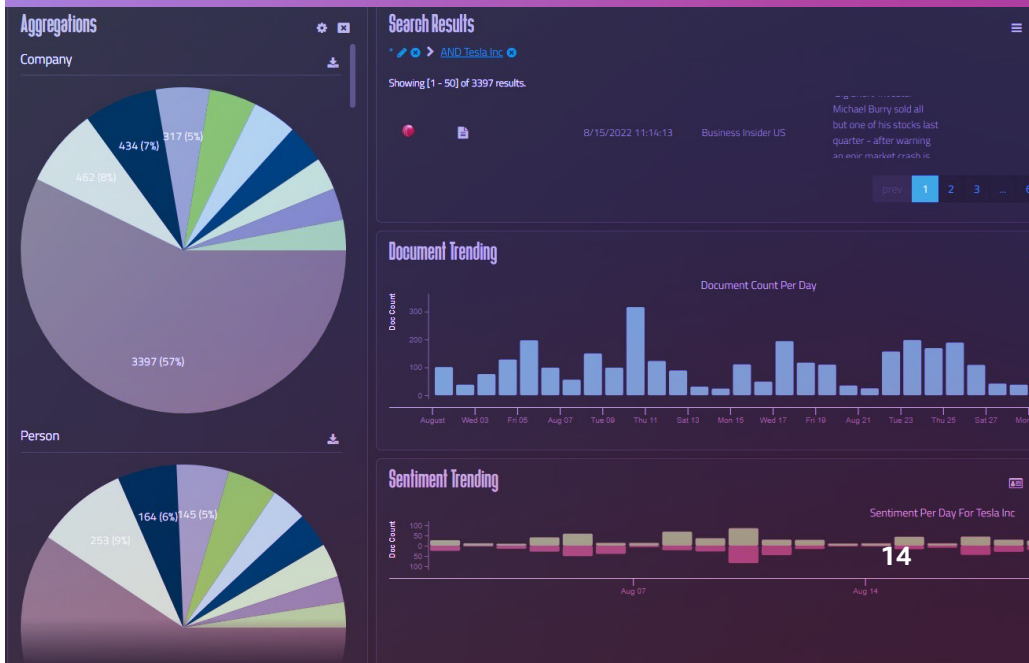
了解详情 ›

亚马逊云科技

FinchAI

“我们将许多生产工作负载迁移到 Inf1 实例，与 GPU 相比，成本降低了 80%。现在，我们正在开发更大、更复杂的模型，以便从书面文本中获得更深刻、更具洞察力的含义。Inf2 实例的性能将使我们能够提供比 Inf1 实例更低的延迟和更高的吞吐量。总之，我们正在提高成本效益，提升实时客户体验，并协助我们的客户从他们的数据中挖掘新的洞察信息。”

Franz Weckesser，Finch Computing 首席架构师



4. Amazon CodeWhisperer

更快、更安全地构建应用程序

Amazon CodeWhisperer 有助于开发人员快速安全地编写代码，而无需离开集成式开发环境（IDE）去研究一些东西。CodeWhisperer 理解以自然语言（英文）编写的评论，并可以实时生成多个代码建议，从而提高开发人员的工作效率。该服务直接在 IDE 代码编辑器中建议完整的函数和逻辑代码块（通常最多包含 10-15 行代码）。CodeWhisperer 包括以下好处：

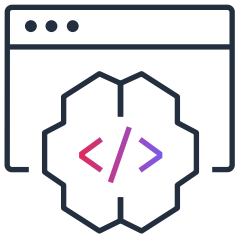
针对亚马逊云科技服务进行了优化
CodeWhisperer 提供针对亚马逊云科技 API 优化的代码建议，使开发人员能够更高效地使用亚马逊云科技服务，包括 Amazon EC2、**Amazon Lambda** 和 Amazon S3。当您在 IDE 中编写代码时，CodeWhisperer 会自动分析您的代码和注释。

内置安全扫描
使用 CodeWhisperer，您可以扫描 Java、JavaScript 和 Python 项目以检测难以发现的漏洞，例如 10 大开放全球应用程序安全项目（OWASP）中的漏洞，或不符合加密库最佳实践和其他类似安全最佳实践的漏洞。该服务会分析 IDE 中的现有代码（无论是由 CodeWhisperer 生成的还是您编写的），高度准确地识别有问题的代码，并提供有关如何修复该代码的智能建议。

负责任地编写代码：开源代码的参考跟踪器
CodeWhisperer 提供了一个内置的参考跟踪器，可以检测代码建议是否类似于开源训练数据，并可以标记此类建议。这些建议用开源项目的存储库 URL、文件引用和许可信息进行了注释，供您在决定是否合并建议的代码之前查看。

负责任地编码：避免偏见
负责任地使用人工智能和机器学习技术是推动持续创新的关键。通过滤除可能被认为有偏见或不公平的代码建议，CodeWhisperer 可协助开发人员避免偏见。

详细了解 CodeWhisperer ›



在一场效率比赛中，使用 Amazon CodeWhisperer 的参与者成功完成任务的可能性比未使用该服务的参与者高 27%，完成任务的速度加快了 57%。

5. 合作的力量：亚马逊云科技合作伙伴 Hugging Face 使用 SageMaker

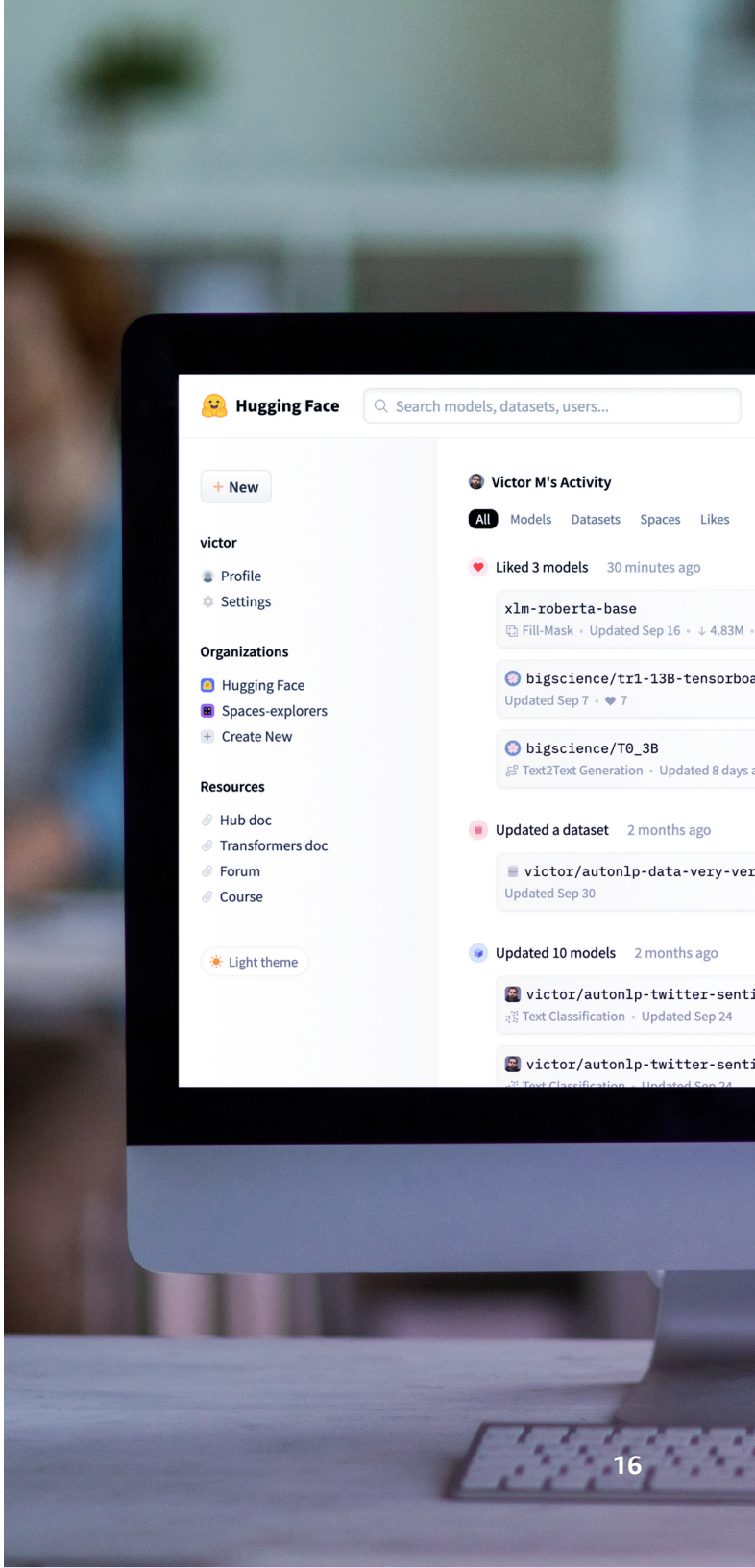
使开源模型更易于访问且更具成本效益

Hugging Face 是一个专注于机器学习的大型开源社区。亚马逊云科技与 Hugging Face 建立了牢固的合作伙伴关系，可加速用于创建生成式人工智能应用程序的大型语言和视觉模型的训练、微调和部署。您可以通过三种方式开始在亚马逊云科技上使用 Hugging Face 模型：通过 Amazon SageMaker JumpStart、Hugging Face 亚马逊云科技深度学习容器（DLC）或将模型部署到 Trainium 或 Inferentia 的教程。Hugging Face DLC 包含优化的转换器、数据集和分词器库，使您能够在数小时（而非数周）内大规模微调和部署生成式人工智能应用程序，并且代码更改极少。

详细了解使用 SageMaker 的 Hugging Face ›

“人工智能的未来就在这里，但分布并不均匀。可访问性和透明度是共享进步和创建工具的关键，有助于实现以明智和负责任的方式使用这些新功能。借助 Amazon SageMaker 和亚马逊云科技设计的芯片，我们的团队和更大的机器学习社区能够将最新研究转化为人人都可以构建的公开可复制模型。”

Clement Delangue，Hugging Face 首席执行官



客户案例

云原生企业正在证实生成式人工智能的潜力

精选客户案例

各种规模的云原生企业都在将生成式人工智能整合到他们的业务中，以便加快创新速度并建立相对于竞争对手的竞争优势。下文介绍了亚马逊云科技如何协助四家云原生企业利用这项革命性技术。

客户案例

InsightFinder 利用亚马逊云科技解决方案成功启动

云原生企业 **InsightFinder** 是一个人工智能驱动的预测可观察性平台，随着使用该平台的学生和教师数量迅速增长，其面临着规模扩展的问题。该公司缺乏内部基础设施来过滤发送的警报。通过将 InsightFinder 引擎与来自 **Amazon CloudWatch** 的数据连接起来，该公司能够快速轻松地获得重要的见解。

参阅案例 ›

InsightFinder

“很多人工智能科技公司认为，您需要在硬件资源上投入大量资金。[而借助亚马逊云科技，] 我们实际上可以以合理的成本，构建一个高性能引擎。”

Helen Gu，InsightFinder 创始人

客户案例

Fraud.net 使用亚马逊云科技机器学习解决方案构建现代反欺诈应用

Fraud.net 是一个欺诈和合规平台，旨在解决给许多贷方、银行、支付处理商和数字商务公司及其客户带来损害的高欺诈率问题。该平台意识到数据缺乏透明度是实现这一目标的最大障碍。Fraud.net 构建了一个可快速部署、可扩展且安全的平台，在该平台上统一欺诈数据并创建切实可行的见解。凭借亚马逊云科技上的事件驱动架构，该云原生企业能够根据事件数量进行扩展和缩减。该公司使用亚马逊云科技解决方案，包括用于计算的 Amazon EC2 和 Lambda 以及用于高度可扩展的对象存储的 Amazon S3。这些解决方案协助其统一和分析三个层面的数据，即客户层面、机构层面和跨机构层面的数据。

参阅案例 ›

亚马逊云科技

Fraud.net

“亚马逊云科技协助我们每秒处理数千笔交易，这一规模在三四年前几乎是不可能实现的。”

Whitney Anderson，Fraud.net 联合创始人兼首席执行官

客户案例

Mantium 在 SageMaker 上使用 DeepSpeed 实现低延迟 GPT-J 推理

Mantium 是一家用于构建和管理 AI 应用程序的全球云平台提供商，使各种规模的企业能够比传统方式更快、更轻松地构建人工智能应用程序和自动化。但是 Mantium 面临一个挑战：开源模型几乎不是为生产级性能而设计的。对于为现代文本生成提供动力的生成式预训练转换器（例如 GPT-J）而言，响应延迟是一个核心障碍。这会使生产部署变得不切实际，甚至不可行。Mantium 利用 DeepSpeed 的推理引擎将优化的 CUDA 内核注入到 Hugging Face 转换器 GPT-J 实现中，显著提高了 GPT-J 的文本生成速度。

参阅案例 ›

MANTIUM

“DeepSpeed 的推理引擎很容易集成到 SageMaker 推理端点。SageMaker 使部署自定义推理端点变得非常容易，并且集成 DeepSpeed 就像添加依赖项和编写几行代码一样简单。”

Joe Hoover，Mantium 公司 R&D 部门的 Senior Applied Scientist

客户案例

Stability AI 借助 SageMaker 获得了弹性、性能并节省了成本

FM 是能适应语言、图像、音频和视频等领域各种下游任务的大型模型，很难训练，因为这些模型需要具有数千个 GPU 或 Trainium 芯片的高性能计算集群，以及可用于高效利用集群的软件。**Stability AI** 是一家开发突破性技术的社区驱动型开源人工智能公司。该公司选择亚马逊云科技作为首选云提供商，提供公有云中有史以来最大的 GPU 集群之一。凭借 SageMaker 托管的基础设施和优化库，Stability AI 的模型训练变得更具弹性、性能更高且更具成本效益，将训练时间和成本减少了一半以上。

参阅案例 ›

stability.ai

“在跨模式扩展我们的开源基础模型方面，亚马逊云科技一直是不可或缺的合作伙伴，我们很高兴将这些模型引入 SageMaker，让数万名开发人员和数百万用户能够充分利用它们。”

Emad Mostaque，Stability AI 创始人兼首席执行官

客户案例

Runway 使用亚马逊云科技扩展了内部研究基础设施

Runway 与亚马逊云科技合作扩展其高性能计算（HPC）集群，并利用我们的研究基础设施在其 Generative Suite 中提供一流的用户体验。Runway 的 Gen-2 系统基于亚马逊云科技进行训练，可以生成带有文本、图像或视频剪辑的新颖视频。Gen-2 改进了 Runway 的多模式生成式模型，代表了用于视频生成的最先进人工智能系统的重大进步。

参阅案例 ›

runway

“在开发和训练这种开创性的视频生成模型方面，亚马逊云科技发挥了重要作用。我们期待着继续共同开创生成式人工智能的无限可能。”

Cristóbal Valenzuela，Runway 联合创始人兼首席执行官

后续步骤

开始使用生成式人工智能

生成式人工智能有望成为几个时代最具颠覆性的技术之一，这种技术可以增强人类创造力、突破创新极限并最大化产出。亚马逊云科技处于最前沿，致力于开发公平、准确的人工智能与机器学习服务，并为您的云原生企业提供必要的工具和指导，帮助您以负责任的方式，构建人工智能与机器学习应用程序。

您的云原生企业是时候行动起来了。您和您的团队熟悉了生成式人工智能的潜力和初始概念后，您就可以开始明确定义自己的目标了。确定具体的实际用例有助于将初始实验保持在更小的范围内，并实现更明确的目标。在考虑数据可用性和质量、选择最适合您的应用的 FM，以及制定实施计划时，建议您与专家合作。生成式人工智能可能具有伦理意义，应在您的用例中予以讨论或解决。

考虑到组织内生成式人工智能的规模和增长，基础设施不应等到事后再考虑；这会对您的成本、规模和能源消耗产生深远影响。与亚马逊云科技的专家合作，可以让您在所有环节和决策中抢占先机。

详细了解利用亚马逊云科技为云原生企业提供的生成式人工智能 ›