亚马逊云科技

中国软件企业云上增长实战指南

第四卷:AI驱动

从功能到基石的生成式AI应用与增长

Contents.

《中国软件企业云上增长实战指南》第四卷—AI驱动

从功能到基石的生成式AI应用与增长

Contents.

《中国软件企业云上增长实战指南》第四卷—AI驱动

从功能到基石的生成式AI应用与增长

ABSTRACT

摘要

生成式人工智能正从一个令人兴奋的技术概念，迅速演变为重塑软件行业格局的核心驱动力。对于中国的软件企业而言，这既是前所未有的机遇，也是不容忽视的挑战。本白皮书是“中国软件企业云上转型实战系列”的第四篇，专为那些希望将生成式AI从“锦上添花”的功能，升级为产品核心基石的软件企业决策者、产品负责人和架构师而设计。我们将深入探讨“为什么是现在”必须拥抱生成式AI的时代背景，提供一套从战略规划到技术选型的系统性框架，剖析高价值的SaaS集成场景与商业模式，并最终聚焦于如何在创新的同时，构建负责任且值得信赖的AI应用，赢得客户的持久信任。这不仅是一份技术指南，更是一份帮助您在AI时代构建持续竞争力的战略蓝图。

NO.1

生成式AI拐点：为何是软件企业的“必答题”？

在过去，将人工智能（AI）集成到软件产品中，通常意味着组建昂贵的数据科学团队，进行漫长而复杂的数据标注和模型训练。这是一个高投入、高门槛的领域，让许多软件企业望而却步。然而，随着基础模型（Foundation Models）的出现和“Transformer”架构的革命性突破，游戏规则被彻底改变了。

我们正处在一个关键的转折点。生成式AI的浪潮，已不再是遥远的趋势，而是拍打在每个软件企业门前的现实。理解“为什么是现在”必须拥抱这股浪潮，是制定未来十年产品战略的基石。

1.1 关键驱动力：从“高不可攀”到“触手可及”

推动这一变革的核心驱动力主要有三点：

基础模型的普及

以GPT系列、Llama系列为代表的基础模型，经过海量数据的预训练，具备了强大的通用能力。这意味着软件企业无需从零开始训练模型，可以直接利用这些模型的“智慧底座”，通过API调用、提示词工程（Prompt Engineering）或轻量级的微调，快速将AI能力集成到产品中。这极大地降低了AI的应用门槛。

“Transformer”架构的革命

2017年提出的Transformer架构，凭借其并行计算能力和对上下文的深刻理解，从根本上提升了模型的训练效率和性能。它不仅是当今几乎所有主流大语言模型（LLM）的核心，其通用性也使其能够处理文本、图像、代码等多种模态，为软件创新提供了无限可能。

市场预期的重塑

ChatGPT的现象级成功，完成了对市场的“用户教育”。如今，无论是企业客户还是个人用户，都开始期待软件能够更智能、更自然地交互。一个“听得懂人话”的软件，正从加分项变为必需品。对于软件企业而言，这既是压力，更是通过提升客户体验来构建差异化优势的绝佳机会。

1.2 新旧范式对比：软件企业的新机遇

为了更直观地理解这场变革，我们可以对比传统机器学习与基于基础模型的生成式AI在软件开发中的不同范式。

特征	传统机器学习	生成式AI（基于基础模型）
开发成本	高：需要大量标注数据和专家团队进行模型训练	低：利用预训练模型，聚焦于API集成和应用层创新
任务特定性	狭窄：一个模型通常只能解决一个特定问题	泛：一个模型可处理摘要、翻译、问答、生成等多种任务
开发体验	复杂：涉及复杂的MLOps和深厚的统计学知识	友好：以API为中心，提示词工程成为新的核心技能
价值实现时间	长：从数据准备到模型部署，周期漫长	短：可通过提示词快速构建原型，验证商业想法

总而言之，生成式AI的出现，让软件企业可以将更多精力从“如何构建AI”转向“如何应用AI创造价值”。这不仅是一次技术升级，更是一场围绕产品创新、客户体验和商业模式的深刻变革。接下来的章节，我们将探讨如何抓住这一历史性机遇。

NO.2

战略先行：构建您的生成式AI产品路线图

在生成式AI的浪潮中，技术本身固然重要，但脱离商业目标的盲目追随是危险的。对于软件企业而言，成功的关键在于将AI能力与产品战略和商业价值紧密结合。一个清晰的路线图，能帮助您从混乱的实验中找到方向，将投资转化为可持续的竞争优势。

2.1 从“客户问题”出发，倒推技术选型

与任何成功的软件功能一样，生成式AI应用的起点应该是客户的痛点，而非技术本身。在规划路线图时，请首先回答以下三个核心问题：

解决了什么客户问题？

您的AI功能是否能帮助客户提高效率、降低成本、创造新的收入，或是提升他们的终端用户体验？例如，通过自然语言查询替代复杂的报表配置，就是一个典型的效率提升场景。

投资回报率（ROI）如何？

集成AI功能需要成本，包括API调用费用、可能的微调成本以及开发资源投入。您需要评估该功能带来的价值（如新客户签约、客户留存率提升、更高定价层级的吸引力）是否能覆盖这些成本。

技术上是否“适合”？

并非所有问题都适合用生成式AI解决。例如，虽然大语言模型可以进行大规模文本翻译，但针对该场景，使用亚马逊云科技的Amazon Translate等专用AI服务可能在成本和性能上更具优势。评估一项技术是否“适合”，需要综合考量其效果、成本和复杂性。

实践建议：

采用亚马逊著名的“逆向工作法”（Working Backwards）。从最终用户将如何受益出发，撰写一份未来的新闻稿（Press Release）和常见问题解答（FAQ），清晰地定义您的AI功能将带来的价值。这个过程将迫使团队聚焦于客户，而非技术本身。

2.2 模型选型：在“多样性”与“控制力”之间取舍

当确定了应用场景后，下一个关键决策是模型选择。这并非一个简单的技术问题，而是关乎企业未来灵活性、成本结构和数据隐私的战略抉择。

核心战略：拥抱“多模型”世界

技术迭代的速度决定了“赢家通吃”的局面难以出现。今天性能最佳的模型，可能在几个月后就被超越。因此，将您的产品战略与单一模型深度绑定，存在巨大的风险。明智的做法是构建一个灵活的架构，允许您根据不同的任务需求，轻松切换和组合来自不同提供商（包括开源社区）的模型。

Amazon Bedrock等服务的设计理念正是基于此，它提供了一个统一的API接口，背后集成了来自多家领先AI公司以及亚马逊自研的多种模型，让您可以“即插即用”，而无需为每种模型重构应用。

开源 vs. 专有：一个战略权衡

考量维度	专有模型（如Claude,GPT-4）	开源模型（如Llama,Falcon）
性能与易用性	通常在通用能力上保持领先，开箱即用，API稳定	性能追赶迅速，需要一定的技术能力进行部署和优化
成本	按需付费（通常基于Token用量），大规模使用时成本较高	部署和推理成本可控，但需要投入硬件和运维资源
数据隐私与控制	需信任服务商的数据处理政策（亚马逊云科技承诺不使用客户数据训练模型）	模型和数据完全在自有环境中，拥有最高控制权
定制与灵活性	定制能力有限（如轻量级微调）	可进行深度修改和优化，不受供应商限制

我们的建议是：

- *对于希望快速验证想法、对数据控制要求不高的通用场景，可以从专有模型入手。*
- *对于希望构建长期护城河、拥有独特数据集、且对数据隐私和成本有严格要求的核心业务，应将开源模型纳入您的核心战略。*

2.3 三大支柱：支撑您的AI产品战略

在制定具体的产品开发计划时，必须将以下三个支柱作为核心考量：

数据隐私与安全

这是客户信任的基石。您的战略必须明确：客户数据是否会离开您的私有环境？是否会被用于训练第三方模型？在多租户SaaS环境中，如何确保租户之间的数据严格隔离？亚马逊云科技提供了完善的解决方案，如在VPC内部署模型、全程加密数据等，确保您对数据拥有完全的控制权。

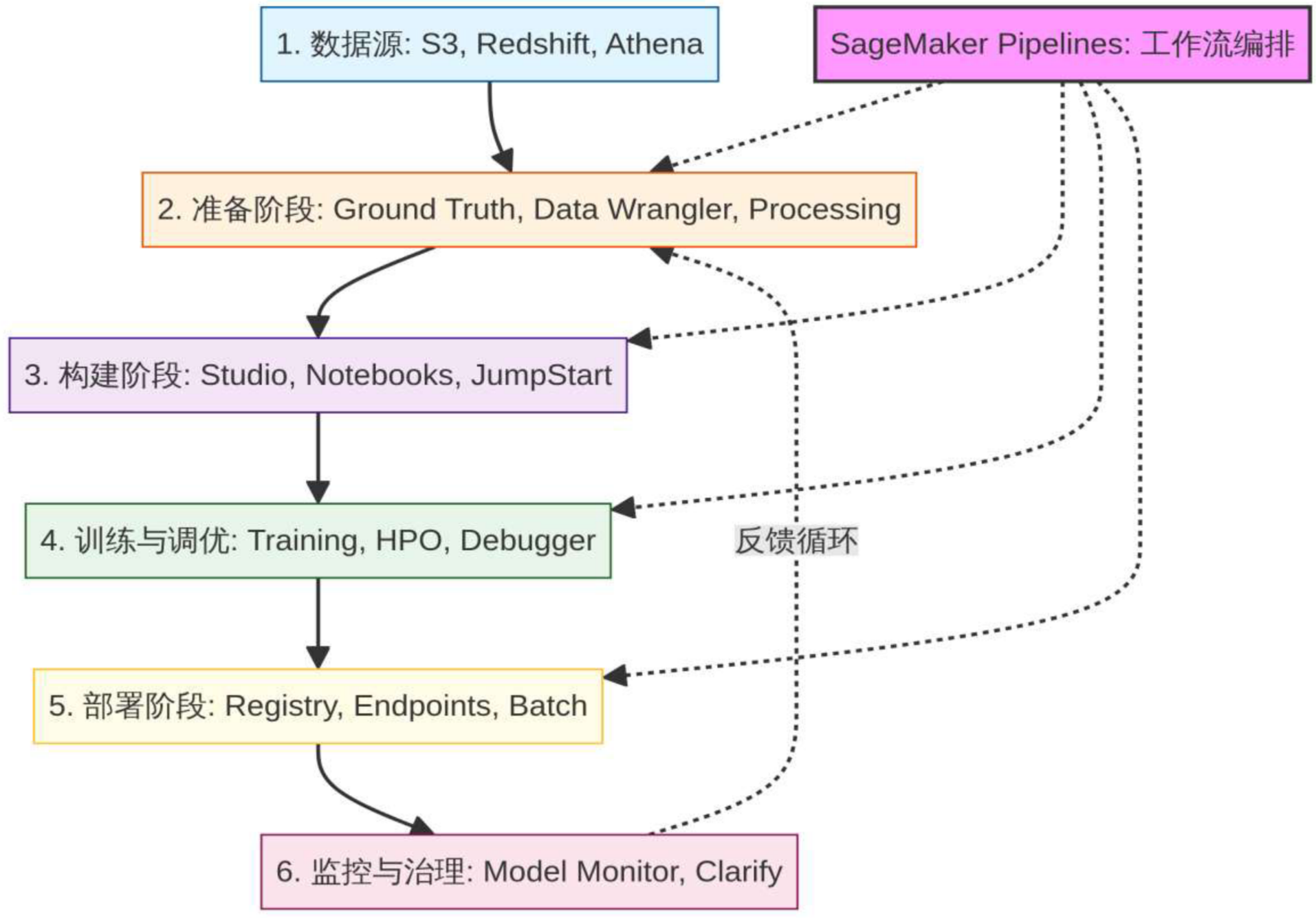
成本效益分析

生成式AI的成本是动态的。您需要建立一个模型来评估不同应用场景的单位成本，并将其与您的定价策略相结合。例如，一个为高级版客户提供的AI功能，其成本是否能被更高的订阅费所覆盖？随着硬件效率提升和模型规模化，成本会持续下降，您的战略也应具备动态调整的能力。

技术栈的灵活性

选择能够支持您从快速实验到规模化部署全过程的技术栈。例如，您可以从Amazon Bedrock这样的无服务器（Serverless）服务开始，快速测试多种模型；当需要更深度的控制和定制时，可以无缝迁移到Amazon SageMaker等平台，进行模型的微调和自管理。这种灵活性是应对技术不确定性的关键。

通过将客户问题、模型选择和三大战略支柱相结合，您将能够构建一个既有远见又切实可行的生成式AI产品路线图，为下一阶段的增长奠定坚实的基础。



Amazon SageMaker核心工作流程图

通过将客户问题、模型选择和三大战略支柱相结合，您将能够构建一个既有远见又切实可行的生成式AI产品路线图，为下一阶段的增长奠定坚实的基础。

NO.3

高价值场景：将生成式AI融入您的SaaS产品

战略规划之后，下一步是落地执行。成功的生成式AI集成，始于识别那些能为客户带来最大价值的具体场景。对于SaaS企业而言，这意味着找到产品与AI能力的最佳结合点，将AI从一个独立的功能，转变为提升核心用户体验的催化剂。以下是三个在软件行业中被反复验证的高价值应用场景。

3.1 场景一：智能副驾（Co-pilot）—— 自然语言查询与数据分析

客户痛点

绝大多数软件产品的核心都围绕着数据，但传统的商业智能（BI）工具和仪表盘往往操作复杂，学习曲线陡峭。业务人员希望能像与同事对话一样，轻松地从数据中获取洞察，而不是学习复杂的查询语言或报表配置。

生成式AI解决方案

在您的产品中嵌入一个“智能副驾”。用户只需用自然语言提问（例如：“展示上季度销售额最高的五个区域”或“分析近期客户流失的主要原因”），AI就能将其自动翻译成数据库查询语句（如SQL），并以图表、摘要或报告的形式呈现结果。这极大地降低了数据分析的门槛，让每个用户都能成为数据分析师。

快速实现路径

您无需从零开始构建这一切。例如，亚马逊云科技的Amazon QuickSight已经发布了基于生成式AI的增强功能，允许您将这种自然语言查询能力直接嵌入到自己的SaaS应用中，为您的客户提供强大的BI体验。

3.2 场景二：主动助手（Proactive Helper）—— 智能客服与应用内支持

客户痛点

用户在使用软件时遇到问题，往往需要中断工作流程，去外部知识库搜索，甚至联系人工客服，体验割裂且效率低下。

生成式AI解决方案

构建一个“主动助手”，它不仅是一个被动的聊天机器人，更是理解用户当前操作上下文的智能向导。

通过结合检索增强生成（Retrieval-Augmented Generation, RAG）技术，这个助手可以访问您私有的、最新的知识库（如产品文档、教程、API参考），为用户提供精准、实时的解答。例如，当用户停留在某个复杂功能的配置页面时，助手可以主动弹出，提供相关设置的说明和最佳实践。

关键技术

RAG模式是实现这一场景的核心。它将基础模型的强大推理能力与您自有知识的准确性相结合，有效避免了模型“胡说八道”的问题。您可以利用LangChain等开源框架，或使用Amazon Bedrock Agents等托管服务，来加速RAG应用的开发。

3.3 场景三：创意加速器（Creative Accelerator）—— 内容生成与流程自动化

客户痛点

许多SaaS产品都涉及重复性的内容创作工作，无论是CRM中的销售邮件撰写、电商平台的产品描述生成，还是低代码平台中的应用逻辑构建，都耗费了用户大量时间。

生成式AI解决方案

将“创意加速器”嵌入到您的工作流中。例如，在CRM中，AI可以根据客户信息和历史互动，自动生成个性化的跟进邮件草稿；在低代码平台中，用户可以用自然语言描述想要实现的功能，AI则自动生成对应的应用组件或代码片段。这不仅提升了效率，更激发了用户的创造力。

开发者赋能

这一理念同样适用于软件开发过程本身。Amazon CodeWhisperer等AI代码生成工具，可以作为开发者的“结对编程伙伴”，实时提供代码建议、发现安全漏洞，显著加速软件的开发和迭代周期。

3.4 如何为AI功能定价：三大商业模式

集成了强大的AI功能后，如何将其商业化是每个产品负责人都需要思考的问题。以下是三种主流的定价策略：

定价模式	描述	适用场景
分层定价 (Tiered Pricing)	将AI功能作为高阶订阅计划的一部分。基础版用户可体验有限的AI功能，而高级版或企业版用户则能解锁全部能力。	适用于希望通过AI功能拉动客户升级、提升客单价的成熟产品。
按用量付费 (Usage-Based Pricing)	基于AI功能的实际使用量进行计费，例如按API调用次数、处理的Token数量或生成的报告份数收费。	适用于AI功能成本较高，且不同客户用量差异巨大的场景，能实现成本与收入的精准匹配。
按价值收费 (Value-Based Pricing)	定价与AI功能为客户创造的直接商业价值挂钩。例如，根据AI帮助客户节约的人力成本或带来的额外收入，按比例收取费用。	这是最理想但也是最难实现的模式，需要与客户建立深度信任，并能清晰地量化AI带来的价值。

选择哪种模式取决于您的产品特性、客户群体和市场策略。通常，一个混合的定价模型（如“分层+超额用量付费”）能提供最佳的灵活性和增长潜力。

NO.4
信任的基石：构建负责任的生成式AI应用

生成式AI的强大能力是一把双刃剑。在追求技术创新的同时，如何负责任地使用AI，保护客户数据，并建立持久的信任，是决定一个SaaS企业能否行稳致远的关键。一个不负责任的AI应用，不仅可能带来法律和声誉风险，更会从根本上侵蚀客户对您产品的信任。

4.1 负责任AI的八大维度

在亚马逊云科技，我们将负责任AI的理念分解为八个关键维度，为软件企业提供了一个全面的治理框架：

公平性 (Fairness)	确保AI系统的输出不会对不同群体产生偏见或歧视。
可解释性 (Explainability)	能够理解并向用户解释AI模型做出特定决策的原因。
隐私 (Privacy)	在AI生命周期的各个阶段，尊重并保护用户的个人信息和数据权利。
安全 (Security)	保护AI系统自身及其处理的数据，免受恶意攻击和滥用。
鲁棒性 (Robustness)	确保AI系统在面对意外输入或复杂情况时，仍能保持稳定和可靠。
治理 (Governance)	建立清晰的流程、角色和责任，确保AI的开发和部署遵循道德和法律规范。
透明度 (Transparency)	让用户清楚地了解他们正在与AI系统互动，并告知他们数据的用途。
真实性 (Authenticity)	努力确保AI生成内容的准确性，并有机制来处理和纠正不准确的信息。

4.2 多租户环境下的安全与隔离：SaaS企业的核心挑战

对于SaaS企业而言，最大的安全挑战莫过于在多租户环境中确保数据的严格隔离。当多个客户共享底层基础设施时，必须从架构层面杜绝任何数据泄露或交叉访问的可能性。在生成式AI的应用中，这一点尤为重要。

核心原则：租户上下文的端到端传递

您的安全策略必须确保，从用户发起请求的那一刻起，代表其身份的“租户上下文”（Tenant Context）就在您的系统中的每一次API调用、每一次数据访问中被严格传递和验证，直到最终的数据存储层。

在生成式AI场景下的具体实践：

模型微调的隔离

当您为不同租户提供定制化的微调模型时，必须确保每个租户只能访问和使用自己的模型。这可以通过IAM（Identity and Access Management）策略来实现，为每个租户的模型资源分配唯一的访问权限。

RAG知识库的隔离

在使用RAG模式时，每个租户的私有知识库必须被严格隔离。如果数据存储在独立的资源中（如不同的S3存储桶），可以使用IAM策略进行访问控制。如果数据存储在共享的数据库中（如池化模型），则需要在应用层面实现更精细的访问逻辑，确保查询时能准确过滤出当前租户的数据。

API调用的安全

所有对Amazon Bedrock等生成式AI服务的API调用，都应在您的后端服务中完成，并使用与租户身份绑定的IAM角色。绝不能将访问模型的密钥直接暴露给前端应用。

4.3 建立客户信任：超越技术的承诺

技术层面的安全保障是基础，但建立客户信任还需要企业做出更广泛的承诺。

透明的AI原则

公开发布您公司的AI使用原则，向客户清晰地阐述您如何确保AI的公平、无偏见和合规。这不仅是品牌建设的一部分，更是对客户的郑重承诺。

数据处理的清晰告知

在服务条款和隐私政策中，用简单易懂的语言告知客户，他们的数据将如何被用于AI功能，以及您采取了哪些措施来保护他们的数据。特别是要明确，客户数据绝不会被用于训练通用的、跨客户的模型。

持续的监控与审计

建立一套持续的监控体系，实时检测AI系统的异常行为和潜在漏洞。定期的安全审计和渗透测试，是验证您安全承诺的必要手段。

最终，客户的信任并非一蹴而就，而是通过每一次负责任的互动、每一次对数据安全的坚定守护，逐步建立起来的。在生成式AI时代，这种信任将是您最宝贵的资产。

NO.5

行动路线图：开启您的生成式AI之旅

理论和场景已经清晰，现在是时候迈出第一步了。我们建议您采用一个循序渐进、从易到难的策略，逐步将生成式AI融入您的产品和组织中。

5.1 三步走策略：从实验到卓越

第一步：快速实验，低成本探索（从零样本推理开始）

目标	验证核心商业假设，找到最有潜力的应用场景。
行动	• 组建一个跨职能的小团队（产品、技术、业务），举办一次内部的生成式AI黑客松。 • 使用Amazon Bedrock等无服务器服务，直接通过API调用预训练的基础模型，进行“零样本”或“少样本”的提示词工程。 • 聚焦于快速构建原型（Prototype），向真实客户演示，收集早期反馈。
产出	1-2个经过初步验证的、具有高潜力的AI功能原型。

第二步：深度集成，打造差异化（拥抱RAG与微调）

目标	将原型转化为正式的产品功能，并利用自有数据构建护城河。
行动	• 针对选定的场景，深入应用RAG技术，将模型与您的私有知识库相结合，提供更精准、更具上下文的回答。 • 如果RAG仍无法满足性能要求，且您拥有高质量的标注数据，可以考虑在Amazon SageMaker上对开源模型进行参数高效微调（PEFT）。 • 设计并实施严格的多租户安全隔离架构。
产出	一个集成到SaaS产品中的、具备差异化优势的AI功能，并配套相应的商业模式。

第三步：规模化运营，持续优化（建立卓越中心）

目标	将AI能力系统性地赋能给整个组织，并建立持续优化的闭环。
行动	• 建立一个AI卓越中心（CoE），负责制定全公司的AI战略、治理规范和最佳实践。 • 投资于人才培养，提升开发团队的提示词工程和AI应用开发能力。 • 建立完善的监控体系（如使用Amazon CloudWatch和Amazon X-Ray），追踪AI功能的性能、成本和用户行为，形成数据驱动的迭代循环。 • 定期使用Amazon Well-Architected框架的SaaS Lens，审视您的AI应用架构，确保其在安全性、可靠性和成本效益方面持续保持卓越。
产出	一个能够持续创新和交付高质量AI应用的组织能力。

5.2 亚马逊云科技：您值得信赖的合作伙伴

无论您处于哪个阶段，亚马逊云科技都致力于为您提供最全面、最灵活、最安全的服务和资源，加速您的生成式AI之旅。

最广泛的模型选择

通过Amazon Bedrock，在一个API背后即可访问多家领先提供商的基础模型，以及亚马逊自研的Titan系列模型。

对开源的坚定支持

在Amazon SageMaker上，您可以轻松部署和微调来自Hugging Face等社区的主流开源模型。

数据优先的理念

我们坚信您的数据只属于您。我们承诺，您在亚马逊云科技上用于生成式AI应用的数据，绝不会被用于训练我们的基础模型。

丰富的赋能资源

从免费的在线课程，到价值1亿美元的生成式AI创新中心，再到与您并肩作战的解决方案架构师和机器学习专家，我们随时准备为您提供支持。

生成式AI的旅程已经开启。现在，就从构建第一个原型开始，将您的软件产品，带入一个全新的智能时代。

附 录

附录A:生成式AI应用场景自查清单

在规划您的AI功能时，可以使用以下清单进行快速评估：

1. **客户价值：**该功能是否能为客户解决一个真实、迫切的问题？
2. **数据基础：**我们是否拥有独特、高质量的数据来支持这个功能，并构建差异化？
3. **技术可行性：**当前的技术（模型性能、相关工具）是否足以支持该功能的实现？
4. **商业模式：**我们是否有清晰的计划，将该功能带来的价值转化为商业收入？
5. **负责任AI：**我们是否已经考虑了该功能可能带来的隐私、安全和公平性风险，并制定了应对策略？

附录

附录B：核心服务与资源

1. Amazon Bedrock：以无服务器方式访问多种领先基础模型的最简单途径。
2. Amazon SageMaker：全托管的机器学习平台，支持从数据准备到模型训练、部署和监控的全流程。
3. Amazon QuickSight：可嵌入的商业智能服务，现已支持生成式AI驱动的自然语言查询。
4. Amazon CodeWhisperer：AI代码伴侣，加速您的软件开发过程。
5. 亚马逊云科技生成式AI创新中心：连接机器学习和AI专家，帮助您构思和构建生成Amazon Well-Architected SaaS Lens：一套针对SaaS工作负载的架构最佳实践，帮助您审视和优化您的应用。