分类号_____　　密级 _____
U D C _____　　编号 _____

兰州财经大学
LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

MBA 学 位 论 文

论文题目 亚马逊机器学习云平台 SageMaker

在中国运营策略优化研究

研 究 生 姓 名：朱煜

指导教师姓名、职称：王学军教授

学科 、专业名称：工商管理

研 究 方 向：创业管理

提 交 日 期：2023.12.17

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：朱煜　　签字日期：2023.12.16

导师签名：王学军　　签字日期：2023.12.10

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，_____（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名：朱煜　　签字日期：2023.12.16

导师签名：王学军　　签字日期：2023.12.10

Research on Operation strategy enhancement for SageMaker in China

Candidate: Zhu Yu

Supervisor: Wang Xuejun

摘　要

随着国家将大数据、人工智能纳入新基建重点发展方向，越来越多的企业希望通过人工智能充分理解自己海量私有数据，提高企业快速决策效率。机器学习云平台基于云计算基础设施，帮助数据科学家、机器学习工程师和企业决策者创建、部署、管理机器学习模型产生商业洞见。

SageMaker 是基于亚马逊云的机器学习平台产品，因其丰富的行业覆盖以及技术前瞻性在世界占据统治地位[1]。SageMaker 于 2020 年四月引入中国，依赖其海外影响力以及众多国际大客户成功故事，本应迅速占领中国市场，然而经过三年的发展，其份额仍在五名以外[2]。本研究以 SageMaker 在中国运营为研究对象，通过深入分析其运营策略提出针对性的优化方案。

本研究以漏斗逻辑，对云计算、机器学习云平台、运营管理的概念进行界定，并对机器学习云平台国内外研究现状进行系统梳理。在明确相关概念和理论基础后，作者展开了对 SageMaker 在中国运营的环境分析。作者首先利用 PEST 宏观环境分析法对其从政治、经济、社会、技术角度加以研究。然后从技术、价格、品牌、生态链四个方面结合对高层的访谈进行运营状况的微观环境分析。最后利用 SWOT 分析法综合宏观和微观分析，发现 SageMaker 在中国运营可以利用的优势、机会以及要克服的劣势、挑战，继而完成了反映其在中国运营策略的波士顿矩阵。在 SageMaker 运营问题的研究中，作者运用访谈法整理出运营存在的主要问题，又分别回归到技术、价格、品牌、生态链四个方面，概括成如下四个核心问题。第一，缺失关键技术功能，导致无法完成平台闭环的技术运营问题。第二，对客户总体拥有成本不友好的价格运营问题。第三，无法大量复制成功案例，做到快速一到一百市场传播的品牌运营问题；第四，生态链发展滞后、存在明显短板的生态链运营问题。在最后的运营策略优化中，作者以 SWOT 分析为基础,以运营管理中“精益生产理论”和“比较优势理论”作为优化依据，综合利用头脑风暴法、鱼骨图法对四个核心运营问题提出优化建议。

关键词：机器学习云平台 运营管理 本土化 亚马逊机器学习平台

Abstract

With the proposal of China's new infrastructure initiative, big data and artificial intelligence have been listed as priority strategic directions. Turning data into business insights through machine learning has become a key development area. Machine learning cloud platforms are an important component aligned with the development of this field. Machine learning cloud platforms, based on cloud computing infrastructure, assist data scientists, machine learning engineers, and software developers in creating, deploying, and managing machine learning models. Machine learning cloud platforms provide a unified development environment, enabling more efficient collaboration across departments and disciplines. Moreover, as machine learning cloud platforms can effectively manage massive amounts of data and accelerate model development by leveraging abundant computational resources, they are more suitable for large and medium-sized enterprise customers.

SageMaker, based on Amazon Web Services, dominates the world in machine learning platforms owing to its extensive industry coverage and technical leadership[1]. Introduced to China in April 2020, SageMaker was expected to swiftly grasp the China market relying on its overseas influence and international customer success stories. However, after three years of development, its market share is outside the top five[2]. This research takes SageMaker's operations in China as the object of study. By

thoroughly analyzing its operational strategy, this research aims to assist SageMaker in better localizing for China. It also hopes to provide references for domestic machine learning cloud platform vendors expanding overseas markets.

Following a funnel logic, this research explains the concepts of cloud computing and machine learning cloud platforms. It also systematically combs the research status of machine learning cloud platforms both domestically and abroad. By introducing lean production theory and comparative advantage theory in operational management, this research provides theoretical basis for researching the operations of SageMaker. Subsequently, the PEST analysis method is utilized to analyze the domestic environment for SageMaker. By drawing a Boston Matrix through SWOT analysis, SageMaker's strengths, opportunities and weaknesses & challenges in China are identified from the perspectives of technological operations, pricing operations, brand operations, and ecosystem operations. By comprehensively applying research methods including expert judgement, brainstorming, and fishbone diagram, the following problems are optimized one by one. Firstly, the lack of key technological functions leads to the inability to complete a closed platform loop. Secondly, it is not cost-friendly regarding customers' total cost of ownership. Thirdly, successful cases cannot be massively replicated to achieve exponential growth in market penetration. Fourthly,

the partner ecosystem is inadequate. Finally, concrete operational optimization suggestions are proposed.

Keyword: AI cloud platform; Operations management; Local operations；AWS SageMaker

目　录

1 绪论

1.1 研究背景和意义

1.1.1 研究背景

随着大数据，人工智能被列入国家级新基建规划，我国逐渐把以大数据为基础的人工智能作为优先发展的重点行业。在企业方面，据德勤 2018 年统计报告，63%的企业正在投入人工智能领域，以期待人工智能可以帮助企业获得行业优势。

伴随着时间的推移，代表着人工智能里程碑级别的应用“ChatGPT”于 2022 年 11 月发布，仅用时六十天注册用户数便破亿，碾压当年社交平台 Facebook 和 Twitter 曾经创下的注册速度纪录。2023 年 2 月 2 日，微软宣布其旗下所有产品全线整合 ChatGPT，并投入 100 亿美元用于 ChatGPT 基础设施的配套投资。ChatGPT 的成功离不开机器学习云平台对它的贡献。正是由于数以万计的算力（计算机计算资源如 CPU，GPU 等）和海量训练数据的支撑，加速了具备多模态能力的大语言模型的诞生。

然而人工智能对于大部分传统企业尚属新生事物，让更多企业用户真正享受到人工智能带来的技术变革需要解决以下三个问题。第一，人工智能需要数据，如何收集、保存海量数据是一个重大挑战；第二，机器学习需要巨量的计算资源，如何多、快、好、省地管理海量计算资源成为第二个挑战；第三，大部分企业对 IT 投入相对有限，甚至没有独立的 IT 部门，也没有专业的算法开发人员，如何低成本的开展机器学习来帮助业务决策，是另一个艰巨挑战。而先天具有完整运营体系的机器学习云平台是这三个问题的解决方案,也是未来开展商用机器学习的主要阵地。

机器学习云平台具有技术门槛高，成本、收益高，生态复杂，注重品牌效应的主要特点。目前只有世界级超大公司才有能力作为辅助业务展开，如亚马逊、谷歌、微软和 IBM。国内也是为数不多的头部互联网企业如阿里巴巴、华为、百度在持续不计成本的投入。可见这个日益增长的市场和以技术为主要驱动力的行业具有高技术壁垒特点。所以急需在其运营管理方面予以足够重视，

控制其技术风险，提高其最大收益，丰富其生态网络，加强品牌知名度才能使产品走向成功。

世界领先的亚马逊机器学习云平台 SageMaker 在中国运营三年。当前中国的数字化程度相对较高，国内机器学习云平台的起步又相对较晚，理应取得比较高的市场占有率，然而事实却不尽如人意，其市场份额始终不能进入前五。这是一个值得研究的现象。然而在查阅资料后，作者在 CCF(中国计算机学会)，ACM（美国计算机协会），IEEE（美国电气电子工程师协会）检索到跟机器学习云平台运营相关的研究非常少。与之形成强烈对比的是自 2017 年后，基于机器学习云平台在各行各业的应用性论文越来越多。因此，作者将机器学习云平台的运营策略优化作为研究对象，以期对提高 SageMaker 在中国运营能力提供借鉴和帮助。

1.1.2 研究意义

首先，本研究立足于机器学习云平台在中国运营落地的策略研究与优化，通过深入分析国内现状，行业竞争格局，定位到 SageMaker 在中国运营的不足，并针对性的提出运营策略优化建议，期望对亚马逊在中国运营机器学习云平台有借鉴作用。也期望对想了解机器学习云平台运营的非技术管理者提供参考信息。

其次，本文在研究 SageMaker 的运营策略过程中利用了比较优势理论跟精益生产理论，对跨国运营机器学习云平台的云厂商提供一个新的研究视角。期待通过借鉴本文提出的分析方法和优化建议帮助其定位不足。

最后，对于计划使用机器学习云平台的企业，也可以参考本文的研究路线通过技术，价格，品牌以及生态链运营的综合评价方式选择适合自身需求的机器学习云平台。

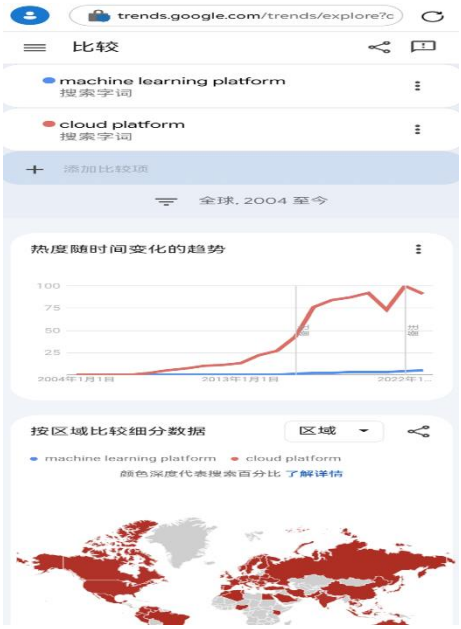
1.2 文献综述

1.2.1 国外研究现状

国外对机器学习云平台运营的探索起步较早，其发展主要经历如下三个阶

段。

探索初期（2002-2009），Buyya 等人在 2009 年发表的论文系统地界定了云计算概念。该学者认为云计算是未来提供大规模算力与数据加工处理的主要平台。需要通过有效的技术运营策略推进产品跟研发的协同发展，其中基于精益生产的小周期快速迭代开发策略可以被广泛借鉴。该研究被视为云计算技术运营领域具有较大影响力的研究成果[3]。谷歌趋势中的红色曲线代表了从 2004 年至 2022 年 12 月全世界互联网用户通过谷歌搜索引擎对关键字“Cloud Platform”的综合统计趋势，图 1.1。该趋势反映出云计算概念从起初的少数人知晓到更多人渴望了解的整体态势，交叉验证了该论文的前瞻性。同一时期 Cusumano 在 2010 年美国计算机协会 ACM 发表了题为“Cloud computing and SaaS as new computing platforms”的研究，认为云计算从技术上可以作为计算平台运营并提供可量化的算力服务，同时也对软件即服务（SaaS）的云服务运营模式做了肯定的评价。该学者认为 SaaS 化是未来技术趋势，短期内也会作为传统软件在部署，维护方面的有利补充[4]。同年 Cusumano 从管理学和战略角度阐述了云计算的诞生对传统软件产品的影响。云平台的持续技术进化直接影响到传统软件从开发到发布以及售后维护的方方面面。从时间维度看，自上世纪 80 年代至 2000 年初，软件的开发环境没有实质的变化，但在 2005 年后的短短 5 年，软件平台从传统数据中心搬到云平台已成趋势。该趋势对云平台厂商和软件企业的技术运营能力都提出了严峻挑战[5]。



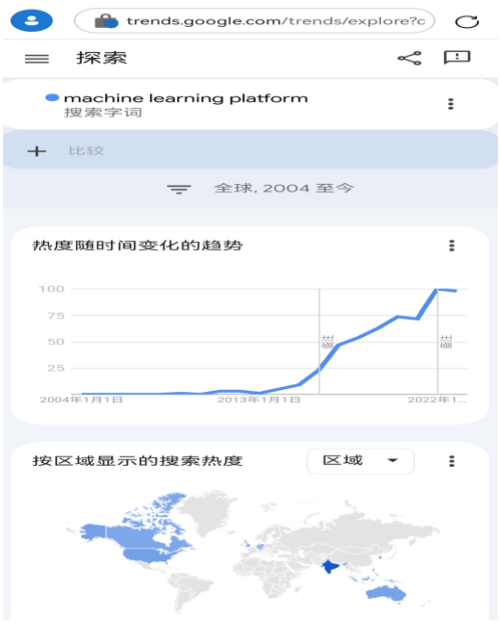


图 1.1 机器学习平台与云平台搜索趋势对比　图 1.2 世界范围机器学习平台搜索趋势

软件即服务（Software as a SaaS）与平台即服务（Platform as a service PaaS）阶段（2010-2017）。随着对云平台认知的普及，越来越多的企业开始在云平台上开展业务，同时也带来了对云平台基础设施运营维护的负担。于是云平台厂商顺势而为，开始了对 SaaS 和 PaaS 产品的研发。该阶段的特点是在基于硬件的基础设施层面上，进一步简化对基础设施的运维操作。Wu 在其“Exploring decisive factors affecting an organization's SaaS adoption: A case study”的研究中指出，虽然云计算和 SaaS 会深刻影响企业软件的开发、发布、维护方式，但不会在短时间内完全取代它。其主要原因是特殊行业对数据的物理安全性以及审计有特殊要求。因此企业不能在短时间将分散的数据迁移至云平台[6]。在 SaaS、PaaS 的成本运营上，Gupta 在其发表的“The usage and adoption of cloud computing by small and medium businesses”研究中指出，价格因素对于中小规模企业尝试云平台以及 SaaS、PaaS 服务有较大影响[7]。Lehmann 在“软件提供商价格策略研究”的研究中认为，对于 SaaS 服务，软件提供商应该按照实际的资源使用量按需收取费用，而不是通过出售软件许可获得收益。这种成本的可负担性使得 SaaS、PaaS 服务被逐渐接受且一直流行至今[8]。图 1.1 中，红色曲线自 2010 年-2017 迅速攀升，体现了基于云平台 SaaS、PaaS 服务在当时的热度。

数据分析与人工智能阶段（2017-今）：大量企业拥抱 SaaS, PaaS 的必然结果是大量的企业数据聚集在云端。而基础算力的多样性使得企业对大数据的商业洞察成为可能。亚马逊电商的推荐系统是基于其存储在云端的大量用户行为数据和云端的海量算力，通过实现个性化推荐机器学习算法而实现的。借鉴该推荐算法的工程化落地经验，亚马逊推出了平台及服务（PaaS）机器学习云平台 SageMaker[9]。Song 在分析如何选择机器学习云平台的研究中“Optimization principles for selecting machine learning platforms”给出了三个建议。第一，目标平台应拥有广泛的生态基础，包括从底层的硬件资源，到上层的数据治理，以及周边的软件配套。第二，要有基于性能的价格优势。第三，要有好的品牌形象，即足够多的可供复制的客户成功案例[10]。Liao 基于 Song 对品牌研究的基础上，在其“Brand equity and customer stickiness in AI cloud services”中更进一步的提出，用户所在行业跟机器学习云平台品牌的粘性存在负相关性。如与亚马逊电商处于竞争关系的其他电商客户采用亚马逊云提供的 AI 服务可能性较小[11]。这也是为什么提供机器学习云平台的科技公司一般都处于各自领域近乎垄断的地位，如微软的 MLOps, 谷歌的 Vertex AI, 亚马逊的 SageMaker、BedRock 等。也正是这些巨头的持续运营跟成功的客户故事，使机器学习云平台自 2017 年被持续关注，如图 1.2 蓝线所示。

1.2.2 国内研究现状

在国家大举发展云计算，大数据，人工智能的新基础设施建设的方针背景下，学界以及企业对机器学习云平台的运营也有很多尝试跟落地成果。由于谷歌无法统计国内搜索趋势，而百度占有国内过半数的文字搜索份额，故以下数据出自 2010 年之后的百度搜索指数。从整体趋势看，我国对云平台的认识起步较晚几乎跟人工智能的认知保持同一节奏，如图 1.3。从积极的方面看，虽然我国发展机器学习云平台起步较晚，但具有在建设云计算平台同时吸取海外机器学习云平台运营经验的后发优势。从发展地域来看，发达的沿海城市以及高校、高科技创业公司富集的省份，直辖市具有较高的研发基础，如图 1.4 所示。

图 1.3 机器学习与云平台搜索趋势对比

图 1.4 国内机器学习搜索热度分布

我国实践机器学习云平台的运营主要经历了如下三个发展阶段。

国外运营经验学习阶段（2002- 2012）：该阶段，我国积极吸收海外云平台在运营方面的先进管理经验，国内各大云厂商相继诞生。2010 年阿里云正式商用，百度、腾讯云于 2011 年几乎同时成立，2012 年华为正式进军公有云市场。随着各大互联网厂商入局公有云，国内学术界也在深入地研究、规划未来在公有云运营方面的发展方向跟战略布局。其代表是 2012 年倪光南院士在中国计算机发表的“面向云计算的服务科学体系”。在其研究中不仅对云平台的服务模式做了总结，还创新性的提出了数据即服务（DataaaS）和安全即服务（SECaaS）的公用云技术运营模式，是对国外流行的 IaaS，PaaS，SaaS 模式的一个补充[12]。

高速发展阶段（2012- 2019）：海外云厂商巨头入华,国内云厂商发展壮大。2013 年 12 月，世界云计算巨头 AWS 宣布进入中国市场在北京提供覆盖全国的本地化云计算服务。2014 年 3 月，微软在中国通过 21 世纪互联展开了本地化云服务业务。虽然国内云厂商起步较晚，但后发优势结合中国本土化运营的经验并没有让海外巨头在中国市场占到统治地位。陈毅在其“Key success factors when Western cloud service providers enter the Chinese market: Case studies

of Microsoft, Amazon, and Salesforce”研究中建议微软，亚马逊提高中文网站质量同时做好良好的客户服务才是其关键的成功因素[13]。在这个阶段，国内云厂商利用海外巨头在中国本土化运营不足的短板，占据了中国云计算市场的前五位。

充分竞争，产品打磨阶段（2019-今）:随着国际化环境在疫情期间骤变，欧美等发达国家对中国芯片实施高压打压政策，导致竞争的进一步白热化。而国内的云厂商只有充分利用其优势才能赢下这生存之战。归纳起来，国内学者有以下几个方面的观点：1.叶劲松认为的技术融合与平台生态系统协同演化可以最大化竞争优势，其论据是技术创新要融入到生态系统中，没有生态的创新会被技术所累[14]。2.陈念主张定制化先行，深入客户场景解决其行业的核心痛点，再复制成功案例到该领域的其他客户[15]。3.张晓等学者认为在当前国际大背景下，随着国外对技术出口限制愈演愈烈，国内巨头的品牌效应凸显，会导致更多的国内企业转向国内云厂商[16]。

1.2.3 国内外研究评述

综合以上国内外对机器学习云平台的研究，可以得出以下几个结论。第一，机器学习云平台是一个新兴的研究领域，以技术驱动为主。美国处于技术领先地位，国内云厂商加速追赶。第二，在机器学习云平台的国内市场中，美国巨头市场份额不高，处于竞争下风。第三，一个机器学习云平台的成功与否必须用一个多维度的评价机制来评估。但是目前并没有一个被广为认可的理论模型和范式可以很好的对其进行评估。

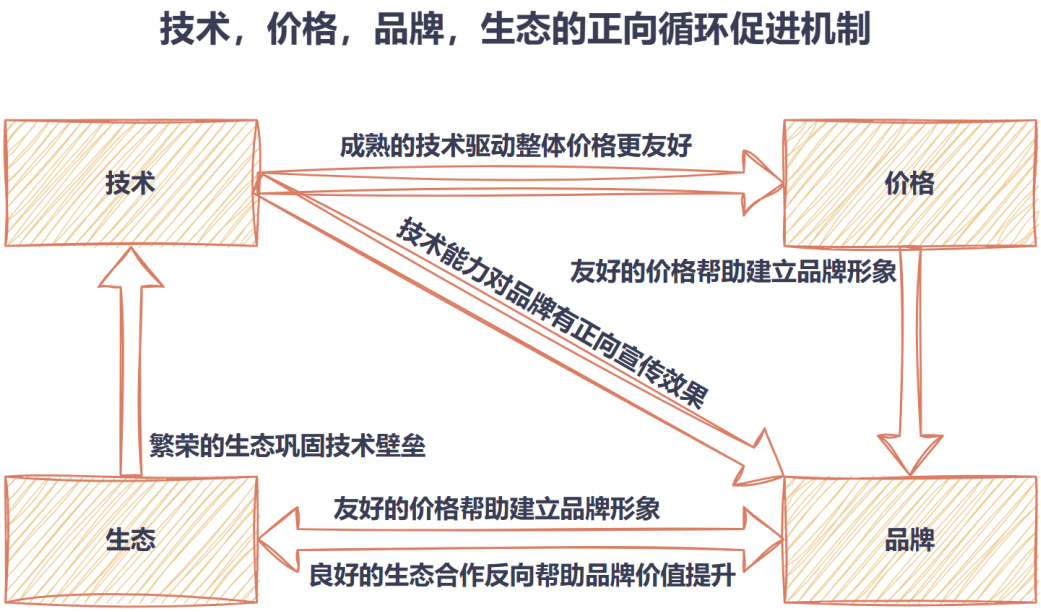


图 1.5 技术，价格，品牌，生态正向循环

以上分析不难看出，从云平台的技术成熟，到按需购买的友好价格模式，促使愿意做技术创新的互联网公司客户首先走上了云端之路，然而可以负担的成本才能让客户在云端之路走的更长久，从而才能积累更多互联网客户的成功故事，继而影响更多非互联网领域的潜在客户。客户的成功故事会正向提高云计算平台的品牌影响力。一个优质的品牌往往会带动一个丰富的生态环境。充分的生态发展又会正向激励技术的迭代创新，形成了一个健康良序的发展闭环，如图 1.5 所示。而以上提及的几个重要方面正是运营管理所涵盖的领域范畴。所以在运营管理的理论指导下，可以很好的平衡各个关系，使其正向促进，达到螺旋式的能力上升的目标。

经过大量的谷歌学术文献检索，以及国内知网、万方等论文期刊的查阅，几乎没有找到机器学习云平台运营策略与优化方面的研究，更多的是如何基于机器学习云平台在具体行业落地人工智能的解决方案和案例研究。而为数不多的与机器学习云平台管理相关的研究也多从流程优化的视角研究如何从技术层面加速其技术实现。

综上所述，利用运营管理理论评估、指导机器学习云平台的运营是一个值得深入研究的课题，也是本文的研究目的。

1.3 研究内容和技术路线

1.3.1 研究内容

本课题以亚马逊机器学习云平台 SageMaker 在中国运营策略的现状分析为起点，针对其在技术，价格，品牌，以及生态链建设的问题，拟完成其运营策略优化的目标。为完成该课题，研究工作具体分为以下四个子问题的研究：

首先，为解决机器学习云平台在中国的技术运营问题，本研究需要进行的工作包括：对中国区 SageMaker 的技术架构进行梳理，根据用户，开源社区，第三方报告对技术问题进行分类。并根据问题分类针对性的进行归因，统计，合并，最终形成可供建议的技术运营优化措施。

其次，对于 SageMaker 在国内价格运营的研究中，作者以内部环境结合外部环境的方式综合分析。在价格的外部环境分析上着重从国家对半导体行业的整体投资情况以及整体利润趋势来做宏观分析。在价格的内部环境分析上，综合统计不同服务以及同一服务不同技术参数的定价策略。最终以优化客户的总体拥有成本为研究目标提出价格运营优化建议。

然后，品牌运营所包括的内容范畴较广，与跨国经营所引起文化差异，媒体环境，用户群体有很大影响。本研究通过头脑风暴法结合鱼骨图法过滤出主要影响因素。聚焦在行业化垂直领域通用解决方案、技术传播和标杆客户案例宣传三个方面，做基于品牌的运营优化。

最后，机器学习云平台的生态链涉及整个云平台和其外围组件。作者通过梳理机器学习云平台的生态体系架构，定义出两类生态体系即机器学习云平台跟自身云平台其他服务之间的生态关系和机器学习云平台跟第三方软件平台之间的生态关系。针对这两类生态进行国内外环境对比分析，最终结合专家访谈对生态运营存在的问题给予建议。

1.3.2 技术路线

基于以上研究内容，本研究采用以下技术线路，如图 1.6 所示。

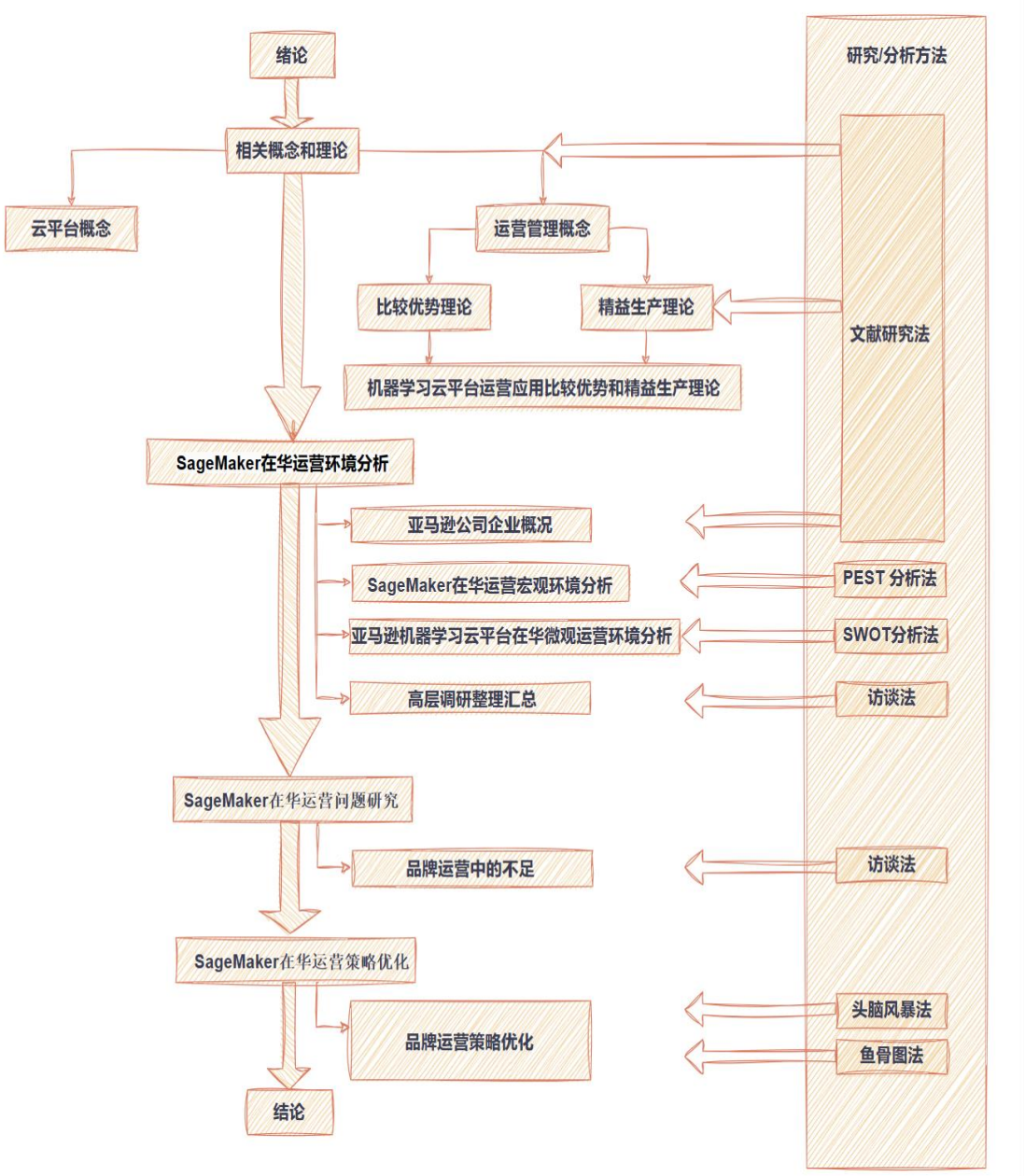


图 1.6 技术线路图

研究的整体章节结构居于上图左侧，以倒树结构自上而下体现，右侧部分为对应章节所用到的研究，分析方法。在相关概念和理论的第二章，主要利用文献分析法对概念加以界定，理论进行说明。在 SageMaker 环境分析章节，主要用到宏观分析法 PEST，SWOT 微观环境分析法，以及高管访谈法对运营环境进行分析。在 SageMaker 运营问题研究中，通过对 SWOT 分析的结果结合高管访谈内容进行概括，总结成四个维度的核心问题。最后在运营策略优化环节，通过头脑风暴法结合鱼骨图法对品牌运营的策略进行优化。

1.4 研究方法

1.4.1 文献研究法

本研究将机器学习云平台和运营管理作为研究对象。通过图书馆借阅大量专业期刊和书籍作为概念，理论的主要来源。此外，研究本着以实事为依据，作者开通了亚马逊云计算的中国以及海外账号，对产品和服务做深入的一手调查，如验证最新中国区 SageMaker 功能上线情况（截至 2022 年 12 月 31 日），不同服务的收费标准以及价格对比等，从而对 SageMaker 国内外机器学习云平台的运营现状有更全面的判断。在研究过程中通过分析、比较、归纳、总结等方法对现有国内外研究进行整理、合并，最终成为本研究的主要分析数据来源。同时对权威第三方咨询机构的行业报告进行研读，如 Gartner 2022 年发布的人工智能开发人员服务魔力象限报告，字母点评 2021 年机器学习云平台采购指南和 2021 年第三季度的 Amazon SageMaker 产品能力评级报告。最后统计亚马逊官方网站的相关数据并做分类汇总，从而基本做到了基于事实的多维度客观分析。

1.4.2 访谈法

研究者与一个或者多个被访问者进行有目的的访谈，由此来了解研究对象所在行业相关的深度见解，从而有目的性地获得有价值的资料。本研究的多个思考如国内 SageMaker 生态链建设的成绩与挑战，SageMaker 在中国品牌运营中采取的本地化策略均来自于亚马逊云科技中国区的高管访谈。对海外 SageMaker 的研究也大量参考了公开的 Re-invent 大会中 AWS 高管的主题演讲，从而获得了可信的一手资料。最终将 SageMaker 中国与海外的运营策略进行了综合归纳，提出了其在中国运营中可以改进的方向以及对应的解决策略。

1.5 分析工具

1.5.1 头脑风暴法

头脑风暴法是一种保证群体决策的创造性，提高决策质量的群体决策方法。在研究机器学习云平台的品牌运营过程中，不同部门对品牌运营中需要重

视跟提高的方向各有侧重。为了充分的暴露所有对品牌运营不利的原因，作者通过组织头脑风暴对所有意见加以收集。

1.5.2 鱼骨图法

因果分析图法又称鱼刺图或树枝图，在研究机器学习云平台的品牌运营过程中，经过头脑风暴环节发现有诸多影响品牌运营的因素，为了聚焦研究重点，通过鱼骨法对所有可能的因素按照从大到小、从粗到细的方法，逐步归纳，找到问题的主因。

1.5.3 PEST 宏观环境分析法

PEST 分析法是一种用于分析宏观环境的战略管理工具[17]，是 Political, Economic, Social 和 Technological 的缩写, 代表了企业宏观环境分析的四个维度。研究者将 PEST 与 SWOT 结合使用，发现 SageMaker 运营的问题，并提出优化方案。PEST 分析法由学者 Francis Aguilar 首次提出, 因其简单易用且视角全面, 已被广泛应用于商业战略管理实践中[18]。

1.5.4 SWOT 分析法

SWOT 即波士顿分析法，也称道斯矩阵，态势分析法。是一种常用的策略性分析工具，用于评估一个项目、组织或个人的优势、劣势、机会和威胁。它帮助识别内部和外部环境的关键因素，为制定战略和做出决策提供指导。本研究通过 SWOT 分析法对亚马逊机器学习云平台 SageMaker 的整体运营状况做微观环境分析，为 SageMaker 运营问题分析以及运营策略优化奠定基础。

2 相关概念和理论基础

2.1 机器学习云平台相关概念

2.1.1 云平台概念

云平台是云计算平台的简称，起初云计算只提供简单的算力和数据存储资源，后来加入了网络，数据库等配套基础设施资源，使其成为一整套面向 IT 服务的平台性解决方案。云平台作为当代信息技术发展的新趋势,正在深刻地影响企业的运营模式,值得从运营管理的学角度研究。云平台的核心是通过互联网实现各种 IT 资源的虚拟化和按需使用。企业用户可以根据自己的需要,弹性调用云服务提供商提供的软件、存储、计算能力等资源[19]。与传统的企业自建数据中心所不同的是,这种使用模式具有高度灵活性和弹性。云平台实现了企业 IT 部门与服务提供方之间的去中间化,企业只需按实际使用量进行支付[20]。按需付费降低了企业的前期投入,实现 IT 资源的按需分配,有助于成本控制。同时,云平台基于分布式架构,可以支持海量用户同时访问,确保服务的高可用性[21]。这大大增强了使用者抵抗运营风险的能力。云平台通过虚拟化和按需使用,实现了 IT 资源的有序安全共享。它可以有效降低企业运营成本,提升运营效率,对于现代企业运用信息技术具有重要管理学意义。

2.1.2 机器学习云平台概念

机器学习云平台作为当今信息技术快速发展的必然产物,正在深刻影响和改变着社会生活及企业组织,有必要从学术研究的视角对其内涵进行明确界定,以奠定后续研究基础。

机器学习云平台是在云平台基础设施之上,集成各类智能算法和模型数据集,面向用户提供人工智能服务的综合性平台[22]。它利用云平台的虚拟化技术和弹性计算能力,使不同规模的用户都可以按需获取机器学习、深度学习、语音识别等人工智能技术服务,避免了用户自行搭建庞大算法环境和模型的高昂成本。与传统手工开发 AI 系统不同,机器学习云平台实现了智能能力的标准化输出和交付。从应用角度看,机器学习云平台降低了用户使用 AI 的门槛,个人和中小企业

无须拥有专门的 AI 技术人员，即可直接使用平台上成熟的算法提升决策质量和工作效率。从组织管理视角看，平台基于海量数据，持续优化模型，并可以利用云计算技术实时获取业务数据，助力企业建立数据驱动的决策体系，实现从经验管理向智能决策的转变[23]。

总体来说，机器学习云平台降低了用户使用 AI 的门槛，支持按需、定制化使用，使智能决策能力普惠化，必然会对企业和社会各领域带来深远影响。

2.2 运营管理概念和运营理论基础

2.2.1 运营管理概念

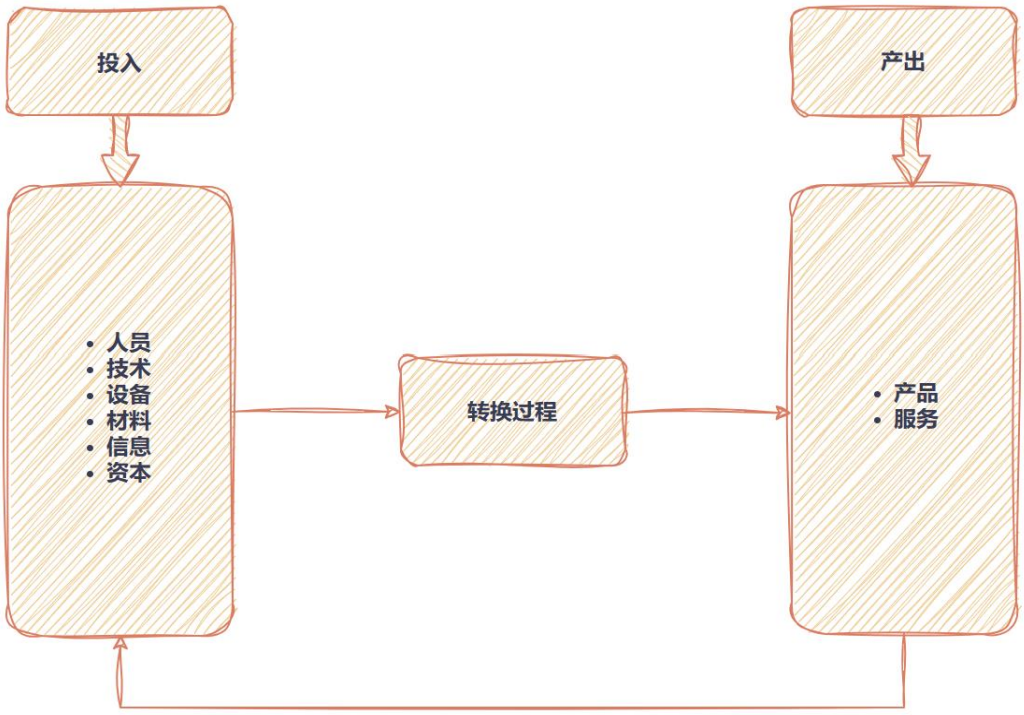


图 2.1 持续运营过程图解

广义上来说运营管理包括生产和运营中的所有行为，例如原材料购买、技术研发、打造品牌、发展生态链等。所谓运营管理，属于管理学范畴，主要包括产品的生产，服务的交付和后期的维护。我们每天在工作生活中都需要很多生活必需品以及服务，这些产品和服务都是在系统的运营管理下有序提供的[24]。运营管理是企业的核心竞争力，是创造利润的引擎，同时也能支撑全球经

济稳步发展，从这一点讲，运营是一个持续迭代的正向循环，如图 2.1 所示。

传统的运营管理经历了几个重要的发展阶段，二十世纪初的“科学管理”阶段、二十世纪三十年代的“质量控制”阶段、二十世纪七十年代的“大量生产”阶段、二十世纪九十年代的“全面质量管理”阶段，基本都是围绕企业本身来进行的，保守且狭隘。随着经济不断地发展和全球化经济时代的到来，新型运营管理模式应运而生，包括运营战略的制定、运营系统设计以及运营系统的运营等多个层次的内容。在实际的运营管理过程中，随着新体系、新技术、新设备等因素的升级，企业也能找到更有效、合理和适合本企业的独特运营管理模式，使运营管理的效果真正达到全面和完善。其中对跨国运营高科技制造业影响深远的运营理论分别是如下介绍的“比较优势理论”跟“精益生产理论”。

2.2.2 比较优势理论

比较优势理论由英国经济学家李嘉图在十九世纪初提出。涉及到跨国运营的主要内容概括为以下几点：1.不同国家生产相同商品存在成本差异。2.每个国家都应该专注生产自己的比较优势商品，以减少社会总成本。3.各国应根据比较优势开展国际贸易，进行商品交换。4.按比较优势理论指导全球贸易,可以使每个国家从贸易中获益。王蒙征在研究中国经济发展理论时通过比较优势理论分析了改革开放以来中国不同地区和产业的比较优势，指出东部地区和出口导向产业具有比较优势。提出发挥比较优势带动经济增长，认为应该充分发挥东部地区的人力和地理优势，发展出口产业[25]。我国电力工业发展不均衡，赵蓓在其题为“比较优势视角下我国电力工业发展战略研究”的研究中,运用比较优势理论分析了我国不同省份在电力工业发展中的比较优势差异。如水力资源优势省份、煤炭资源优势省份等。指出应该根据各省比较优势特点制定电力发展战略，同时建议加强区域电力合作,实现优势互补[26]。以上研究足以证明比较优势理论的普适性，对于全球化运营机器学习云平台的生态链，价格有很大指导意义。

2.2.3 精益生产理论

精益生产理论起源于二十世纪五十年代的丰田生产系统,主要内容有:1.持续消除生产过程中的浪费，如等待时间、过度加工等。2.建立流水线生产,实现连续平稳的作业流。3.实施拉动生产方式,根据下道工序需求及时补充物料。4.推行全员参与,鼓励员工持续提出改进意见。5.与供应商形成战略合作关系,共同提高质量[27]。张爱卿在其论文“基于精益生产理念的供应链管理模式创新研究”中充分利用精益生产理论，提出在供应链系统中应用精益生产的理念即消除供应链环节中的各种浪费,实现流程优化。强调供应链伙伴之间的战略协作,供应商和上下游生产企业之间形成战略合作利益共同体[28]。对于机器学习云平台全球化运营的目标，这种精益的生产合作，最大限度减少浪费，流程的优化，会对技术运营的加速迭代，生态链合作伙伴的快速增长有实际借鉴意义。

2.3 机器学习云平台运营应用比较优势和精益生产理论

比较优势理论和精益生产理论虽然都注重提高效率、降低成本,但角度不同。对于运营新兴的机器学习云平台来说，都有很好的借鉴跟互补意义。将这两个理论充分融合在机器学习云平台的运营中可以持续提高其核心竞争力。其中,技术运营、价格运营、生态链运营和品牌运营尤为关键。

技术运营是指平台提供者通过技术手段提升产品与服务质量的持续优化过程。具体来说，技术运营包括算法迭代、模型优化、平台监控等。比如平台提供方可建立敏捷的技术研发团队,实施快速迭代开发,并采用在线 A/B 测试评估新功能上线等。还需要监控平台稳定性，识别技术风险,以保证服务持续可用，这是精益生产理论的主要内容与方法。

价格运营是根据市场与成本状况制定和调整产品或服务价格的过程。机器学习云平台可根据用户规模和使用时长，采取弹性的计价策略。也可以根据不同客户细分采取差异化定价，实现成本最小化。具体实施上，可以提供按量计费的统一价格,也可以考虑订阅制与团体折扣制组合。持续评估不同模式对不同价格敏感度客户的影响，有助于找到最佳方案。

生态链运营是指平台提供方通过生态链合作伙伴关系构建综合解决方案的

过程。例如与芯片厂商合作定制 AI 专属芯片，与行业客户合作开发定制模型，以及对特殊需求提供差异化服务。同时，与渠道商合作，利用其销售网络进行产品推广。与生态伙伴在生态链的上下游协同创新是获得共同核心竞争力的重要途径。

品牌运营是指平台提供方通过传播手段塑造用户知觉,建立认知印象的过程。明确的品牌定位有助于吸引目标用户群。此外,提供出色的客户服务可以提高品牌忠诚度。同时，关注客户反馈并不断优化产品,是提升品牌美誉度的重要方式。 图 2.1,是融合比较优势理论和精益生产理论在机器学习云平台持续运营体系中的整体架构图。

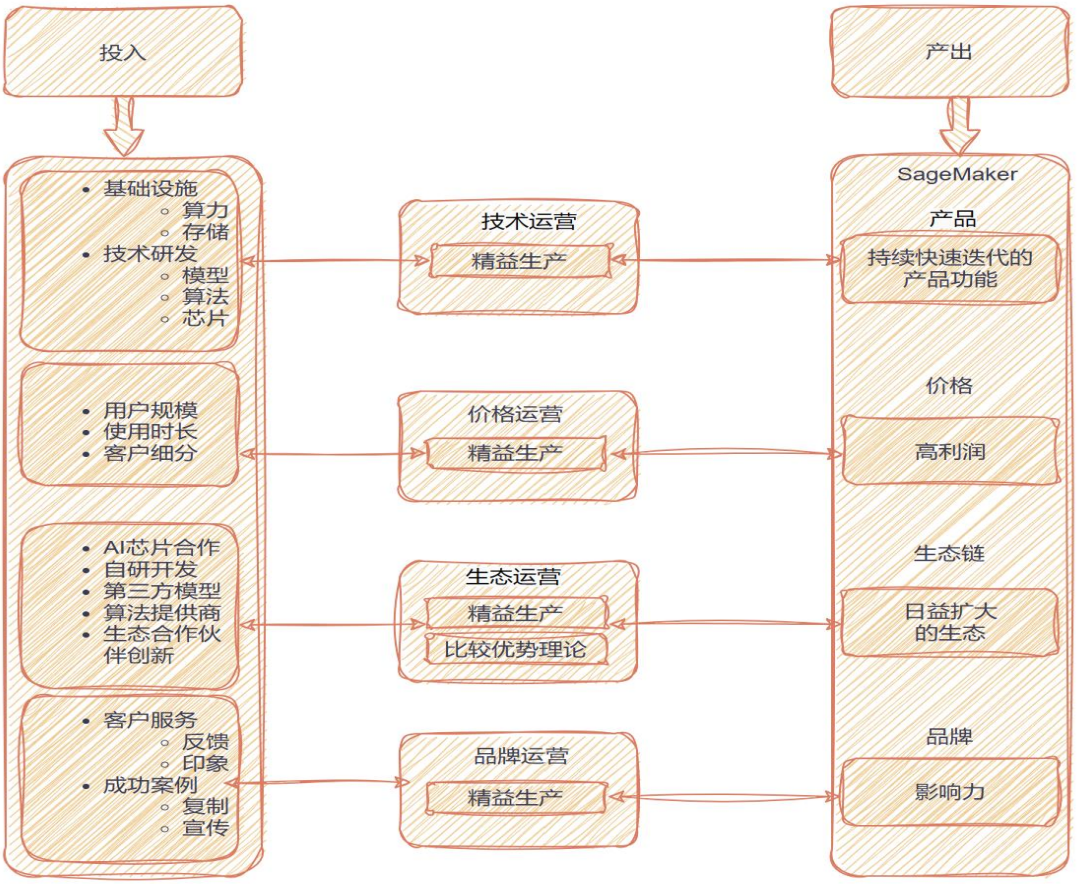


图 2.1 融入精益生产和比较优势理论到机器学习云平台的运营

综上所述,在精益生产理论和比较优势理论指导下，技术运营、价格运营、生态链运营和品牌运营是机器学习云平台实现持续发展的关键运营手段。

3 亚马逊机器学习云平台 SageMaker 在中国运营环境分析

3.1 亚马逊公司概况

3.1.1 发展历程

亚马逊公司成立于 1994 年，是一家总部位于美国华盛顿州的跨国电子商务公司。亚马逊公司的创始人为杰夫·贝索斯（Jeff Bezos），公司最初是一家以在线销售图书为主的公司，现在已经发展成为全球最大的在线零售商之一，其企业的领导力准则也被众多公司认可并效法。

1995 年，亚马逊公司网站正式上线销售图书，并开始筹备上市。1997 年 5 月 15 日，亚马逊公司在纳斯达克股票市场上市，募集资金达 5500 万美元，成就了全球首个电商巨头为互联网的兴起起到了推波助澜的作用。

1999 年，亚马逊公司开始推出更多种类的商品，如音乐、电影和玩具等。次年，亚马逊公司推出了其市场平台，允许第三方卖家销售商品，盘活了更多线下渠道进一步巩固自己在电商行业的领先优地位，体现了其卓越的生态运营能力。亚马逊还首创的提出了黑五抢购等品牌运营策略，成为了当时电商品牌宣传的经典案例，后期纷纷被国内的电商企业效仿。

由于各种平台的接入以及市场促销活动的盛行，亚马逊电商出现了诸多运营问题，其中影响最大的是资源浪费严重。每年圣诞季亚马逊电商都会做规模空前的酬宾大促，会吸引数以千万的顾客在线下单，为了应对订单的高峰，亚马逊电商通常会购买几倍于平时用量的服务器以应对抢购的高峰。然而在促销季之后这些购置的资源无法充分利用造成巨大的浪费。于是 2002 年亚马逊开始内部构思弹性计算服务也就是云平台的前身，将闲置的计算资源通过接口调用的方式转卖出去，实现成本的回收。该资源转卖服务于 2006 年正式对外发布，并命名其为 Elastic Cloud Compute（弹性云计算服务）简称 EC2，开创了云计算服务的先河。自此亚马逊开始了其风光无限的云计算业务-亚马逊云服务（Amazon Web Services），体现了强大的技术创新运营能力。

自 2010 年后，亚马逊公司先收购了在线鞋类零售商 Zappos，扩大了其在线零售业务范围。又推出了 Echo 智能音箱和 Alexa 语音助手，标志着公司进军智

能家居市场。此后，亚马逊公司不断拓展其业务领域，包括引领购物体验创新的无人超市、电子商务支付、食品配送、物流和医疗健康等，覆盖了众多领域的诸多行业。

亚马逊公司的主要业务包括：亚马逊在线商店、亚马逊云服务、数字媒体（包括电子书、电影、音乐和游戏等）、电子设备（如亚马逊 Kindle 电子书阅读器、亚马逊 Echo 智能音箱等）、以及其他零售业务。伴随着其在不同领域的创新与持续优化运营，亚马逊公司的市值在 2019 年达到了全球第一。同时也是全球最大的在线零售商和云服务提供商。截至 2021 年 9 月底，亚马逊公司的全球员工人数已经超过 140 万人，其总营收为 3867 亿美元，净利润为 214 亿美元。

3.1.2 亚马逊云服务（AWS）运营概况

AWS (Amazon Web Services) 是亚马逊公司的云计算服务部门，提供云存储、计算、数据库、分析、人工智能等各种云服务，其中 SageMaker 是 AWS 提供的机器学习云平台产品。AWS 一直以来都是亚马逊利润的主要来源之一。AWS 的客户数量庞大，包括了全球各种规模的企业、政府机构、学术机构、初创企业等。AWS 的客户数量和规模也在不断增长。AWS 提供超过两百大类云服务，覆盖计算、存储、网络、安全、数据库、数据分析、人工智能、机器学习、物联网、混合云等，甚至涉及还处于前瞻性物理实验阶段的量子计算和卫星数据服务。亚马逊云科技（Amazon Web Services 在中国运营的注册商标，海外简称 AWS）现已在全球三十个地区运营九十六个可用区（Availability Zone），超过四百个边缘站点和十三个区域性边缘缓存站点，服务超过两百四十个国家和区域。亚马逊云科技在全球拥有数百万活跃客户和数万个生态链合作伙伴，拥有庞大且具活力的生态系统。几乎所有行业和规模的客户（包括初创公司、企业和公共部门组织）都可以在亚马逊云上找到可供借鉴的使用案例。从产品服务而言，亚马逊云科技提供从基础设施及服务（IAAS），平台即服务（PASS），软件即服务（SAAS）的综合 IT 技术解决方案。从客户的角度来说，利用亚马逊云科技提供的各项服务可以实现安全、高效、高性能、高弹性的基础设施。同时亚马逊云科技还有成熟的方法论帮助构建高性价比、卓越运

营、可持续发展的业务平台和工作负载。亚马逊云科技于 2014 年正式进入中国，目前运营了两个区域分别位于北京和宁夏。本文之后提到的亚马逊云科技专指中国区的亚马逊云服务。

目前只有亚马逊云科技在其财报中体现了云平台的盈利能力，其他主流云厂商均没有对云平台的盈利能力的数据支持。对于中国市场而言，亚马逊云科技处于市场的主要领导者地位。2022 年 9 月 26 日，中国 ICT 产业权威的市场研究和咨询机构计世资讯（CCW Research）发布《2021-2022 年中国公有云市场现状及趋势研究报告》（以下简称《报告》），亚马逊云科技被评为“2021 年中国公有云 IaaS 市场领导者”，其创新发展能力在报告中居于首位[29]，如图 3.1 所示。报告指出，2021 年国内公有云（包括 IaaS、PaaS、SaaS）规模约为 1515.1 亿元，增速为 34.6%；预计 2022 年公有云规模增速将降至 23.8%，整体规模约为 1875.6 亿元；2023 年增速将提升至 26.7%，整体规模达到 2376.4 亿元。在这样的市场规模背景下，有必要对领导者的运营策略做深入分析。

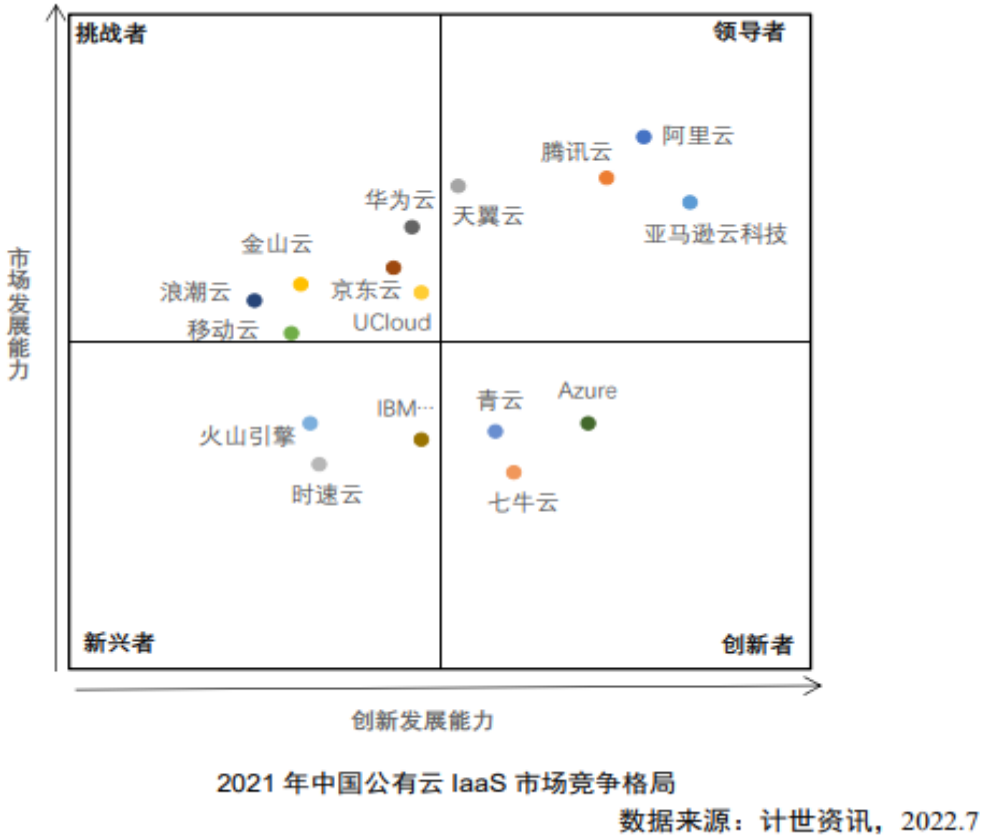


图 3.1 2021 年公有云市场(IaaS)市场竞争格局

上图 3.1 是中国 ICT 的分析结果，跟 Van Baker[30]中对全球权威 IT 咨询企业 Gartner 如图 3.2 的分析有近乎一致的结论。

图 3.2 Gartner 2022 年人工智能平台能力象限图

结合以上国内、国际的能力象限分析，可以归纳为云平台的运营能力主要体现在以下几个方面。第一，技术运营能力，体现了能够对客户提供充分的场景支持能力。第二，价格运营能力，尤其是统筹国内国外的市场价格，其中牵扯到汇率波动，人力资源定价等。第三，品牌运营能力，尤其对亚马逊这样的国际公司，如何迅速本土化，成为本土客户手口相传的优秀品牌。第四，生态链运营，对于目前部分国家的逆全球化而导致的数据，芯片的进出口限制对产业链运营能力的挑战。以上所列的任何一个运营能力也会彼此促进反之也会互相掣肘。

3.2 亚马逊机器学习云平台在中国 PEST 宏观环境分析

为了全面对中国机器学习云平台的整体宏观状况进行分析，研究者使用 PEST 分析方法，从以下四个维度进行分析。

政治因素（Political）：国内政治气氛对发展机器学习和人工智能产业非常鼓励，同时也在监管上有配套的政策法规。如工信部印发《新型数据中心发展三年行动计划（2021-2023 年）》，计划用三年时间，基本形成布局合理、技术先进、绿色低碳、算力规模与数字经济增长相适应的新型数据中心发展格局。大力推进数据建设，人工智能建设，这是对机器学习云平台的重大利好政策。在监管侧，2022 年发布的《互联网信息服务算法推荐管理规定》，2023 年发布的《互联网信息服务深度合成管理规定》，以及 2023 年 8 月 15 日实施的《生成式人工智能服务管理暂行办法》，体现了国家对规范发展机器学习及其平台所持的鼓励并严格监管的态度。

经济因素（Economic）：疫情期间全球经济陷入整体发展停滞阶段，美元走强，导致国内出口受阻，我国尚未确立经济回弹趋势。随着美商务部 2022 年 10 月对我国先进半导体工艺 AI 芯片实施出口限制，我国机器学习行业因此受到巨大冲击。基于 AI 芯片的机器学习平台的影响尤其严重。生成式 AI 所依赖的算力主要基于 0.5 纳米工艺的 AI 芯片。高制程芯片不足已经限制了云计算行业在算力方面的发展，造成了不同程度的算力不足。为了改善该被动局面，国家对 AI 芯片产业进行大规模扶持。来自国家工业和信息化部的统计，自 2022 年 1 月至 2023 年 6 月，我国电子信息制造业的利润总额持续变少，如图 3.3 所示，而累计投资额一直维持在 9%以上的增幅。以上统计数据正反应了国家正在积极引导、寻求突破，如图 3.4 所示。随着搭载了国产芯片的华为 Mate60 发布，未来国产芯片的前景被一致看好，也为机器学习云平台在算力的国产化替代提供了坚实基础。

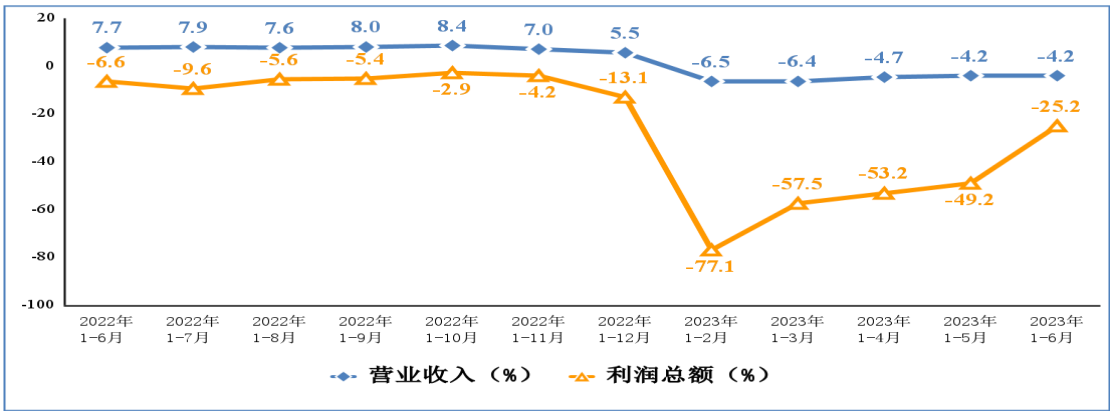


图 3.3 利润总额累计增速

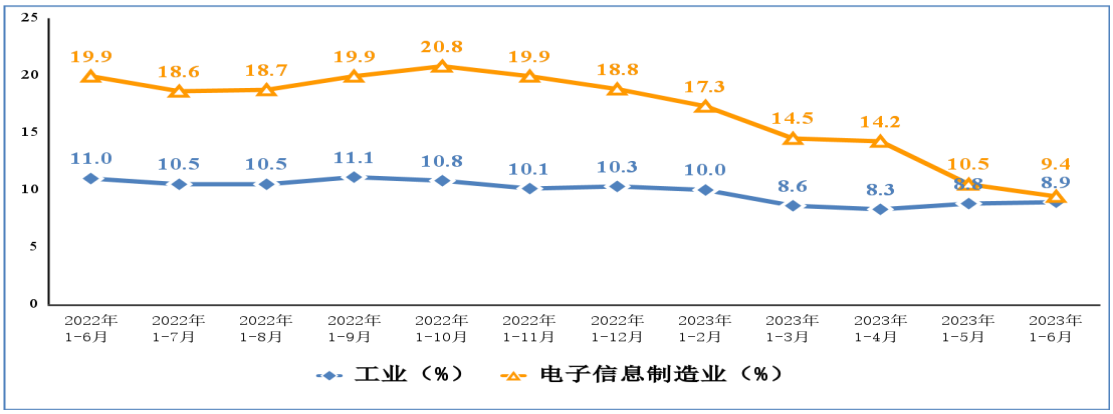


图 3.4 固定资产投资累计增速

社会因素（Social）：机器学习云平台的开发与创新离不开高质量的研发人员，算法工程师，训练数据标注员和数据分析师。目前国内从事该行业的人数仍有较大缺口。但是目前全国理工类高校基本开设了大数据，人工智能专业，未来 3-5 年，这方面的人才供应会有较大改善。从社会认知层面看，大众已享受到人工智能带来的便利，如人脸识别，智能语音助手等，而且变得越来越依赖这些科技带来的整体效率提升。这些功能都是通过机器学云平台赋能给相应产品和服务的。疫情期间机器学习云平台所发挥的积极作用也有目共睹，如基于云平台研发的无人驾驶车辆给隔离区配送物资，又或是健康宝的实时智能数据分析服务在疫情期间发挥了巨大的风控隔离指导作用。

技术因素(Technological)：生成式 AI 在 2022 年底的兴起带动了大量新场景应用。如基于文字生成图片的功能，可以在社交类 APP，游戏人物设计中得到应用。在基于文字生成回答的内容生成领域，基于大语言模型的智能问答系统得到广泛应用。这些应用极大提高了人们生成内容和检索信息的效率。但是新的技术也带来了其他的挑战。例如截至 2023 年 8 月，这些生成式技术所依赖的机器学习云平台并没有直接的功能可以帮助模型训练方对模型的输出从道德规范，政治正确上加以约束。这也是目前机器学习云平台厂商在努力寻求解决的技术难点。

3.3 亚马逊机器学习云平台在中国微观运营环境分析

3.3.1 SageMaker 产品构成

为解决机器学习本身存在的诸多挑战，让数据科学家、算法工程师、业务开发者都能轻松驾驭机器学习，亚马逊公司于 2017 年 11 月推出了 Amazon SageMaker。经历 5 年磨砺，Amazon SageMaker 进化成一套完全托管的机器学习平台服务，可以帮助数据科学家和开发人员快速轻松地构建、训练和部署任何规模的机器学习模型，而无需关注底层资源的管理和运维工作。

Amazon SageMaker 作为一个工具集，提供用于机器学习端到端的所有组件，包括数据标记、数据处理、算法设计、模型训练、训练调试、超参调优、模型部署、模型监控等，使得机器学习变得更为简单和轻松；同时，SageMaker 依托于 AWS 强大的底层资源，提供了高性能 CPU、GPU、弹性推理加速卡等丰富的计算资源和充足的算力，使得模型研发和部署更为轻松和高效。

机器学习训练任务的核心在于选择包含算法和框架在内的容器镜像，关于训练镜像的来源，用户既可以选择由 SageMaker 提供的多种内置算法镜像，也可以选择基于 SageMaker 内置框架（TensorFlow、Apache MXNet、PyTorch、Scikit-learn、XGBoost）镜像，结合用户自己的代码来完成训练。如果用户想使用自己的训练代码或自己研发的软件框架来完成模型的训练，也可以通过 SageMaker 的容器方式来实现。SageMaker 总体架构图如下 3.6 所示：

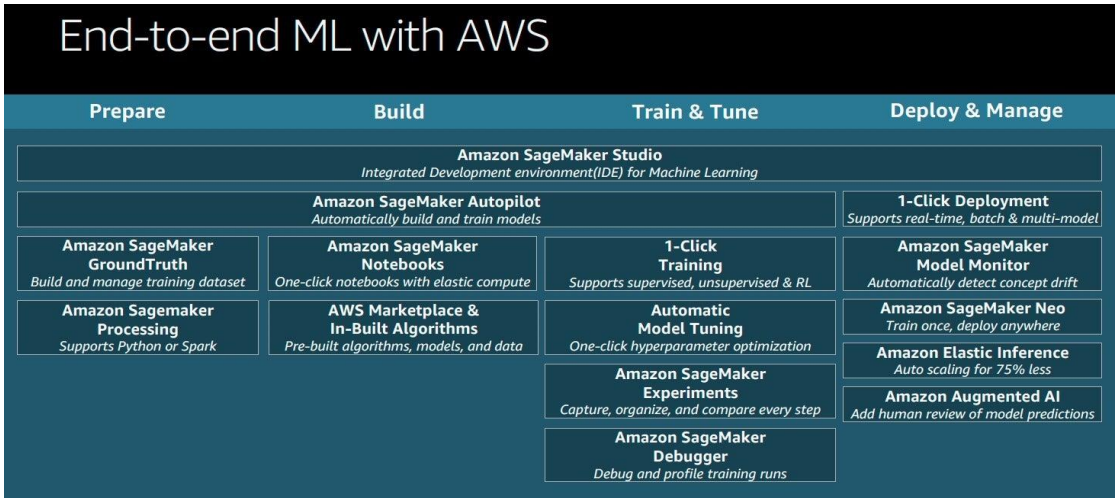


图 3.6 SageMaker 产品介绍 - 源自 SageMaker 官网

亚马逊机器学习云平台 SageMaker 被权威第三方 IT 评估机构 Gartner 认定为机器学习云平台的领导者,其首席调研专家 Van Baker[30]指出，亚马逊云科技是这个魔力象限中的领导者。其人工智能服务，包括 Amazon SageMaker 和其他流行的语言和视觉服务，旨在自动化整个人工智能开发和运营周期。它在机器学习云平台的开发者服务（AIDS Cloud AI developer services）市场拥有强大的全球影响力，客户遍布几乎所有行业。SageMaker 允许客户在他们专业人员协助下或在咨询生态链合作伙伴的帮助下自行构建解决方案。亚马逊云科技的 SageMaker 是商业化应用机器学习有吸引力的选择，因为它的运营成本低，而且 AI 服务和基础设施选择范围广，可供借鉴的成熟方案多。Barry[1]也在其发表于 2022 年 7 月的电子调研报告中列出了亚马逊机器学习云平台的主要优势也体现在覆盖场景丰富，以及开箱即用的特点。下图 3.7 是 SageMaker 在机器学习产品在各个机器学习阶段的映射关系。之所以有这样的整体评价也反映出 SageMaker 在技术创新，价格策略丰富，生态链完整和注重品牌运营的多元运营能力。

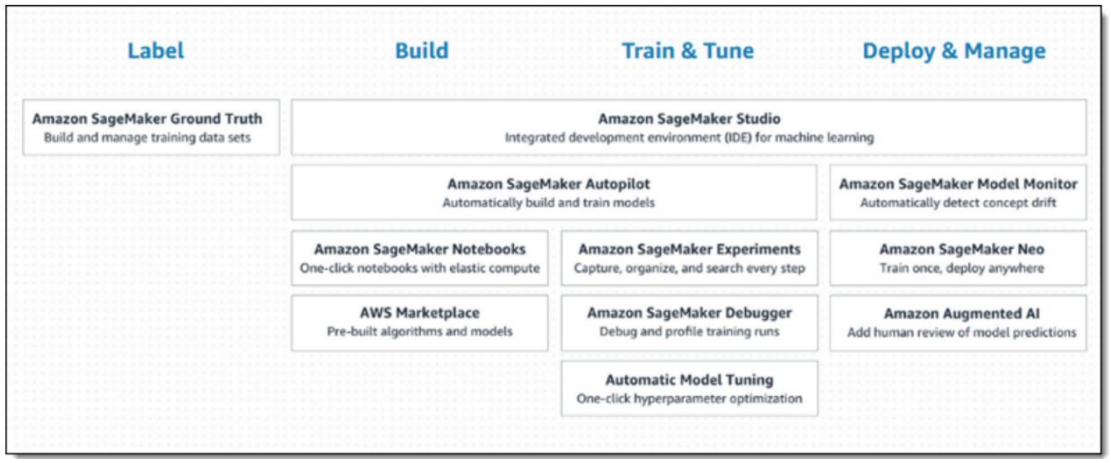


图 3.7 SageMaker 功能覆盖-源自 SageMaker 官网

3.3.2 SageMaker 技术运营现状分析

技术运营，是通过自身的技术能力生产出产品或服务，利用其产品或服务帮助其客户解决业务问题的。对于亚马逊机器学习云平台 SageMaker 来说，其技术运营能力主要表现在其产品的新功能上线速度方面，以及与其他产品的融合方面。

从平台软件更新迭代看，2022 年 re:Invent 全球大会上，SageMaker 推出八项新功能。其功能可以使众多开发人员、数据科学家和业务分析师受益。使用 SageMaker 提供的全托管基础设施、工具和工作流，轻松快速地构建、训练和部署机器学习模型。对于与其他云服务的集成来说，SageMaker 可以和运算服务如 EC2、Lambda 以及面向容器的 ECR 服务做无缝的衔接，还可以与 S3，EFS,EBS 等数据服务集成。如下图 3.8 所示，SageMaker 作为上层的全托管的机器学习云平台，可以适配多种机器学习框架，基础计算服务以及存储服务：

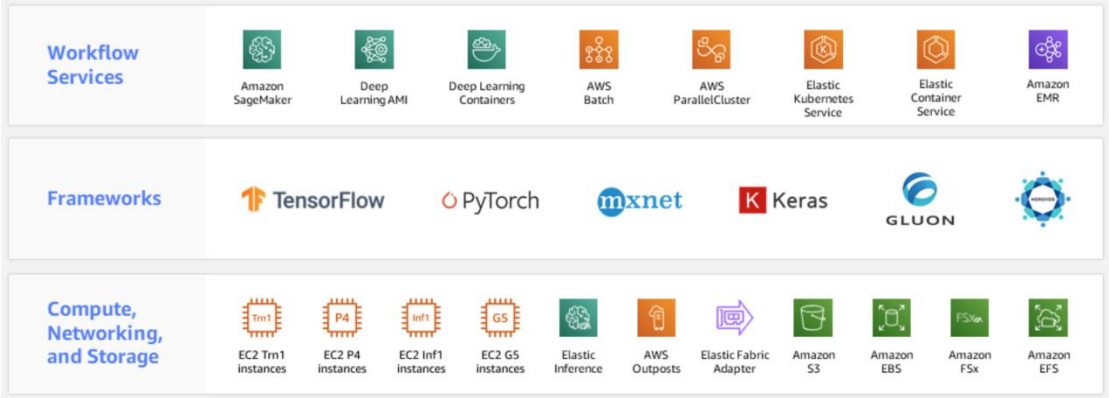


图 3.8 机器学习生态框架

从基础设施的推陈出新看，截至 2022 年 12 月,可以为 SageMaker 提供算力的计算资源有：1. Amazon EC2 Trn1 实例,由 AWS Trainium 芯片提供支持，专为高性能深度学习而构建，并为在云中训练深度学习模型提供高性价比。2. Amazon EC2 DL1 实例由英特尔公司 Habana Labs 的 Gaudi 加速器提供支持，旨在训练深度学习模型。3. P4d 实例利用 8 个 NVIDIA A100 Tensor Core GPU、400 Gbps 实例联网能力以及 NVIDIA GPU Direct RDMA（远程直接内存存取）对 Elastic Fabric Adapter(EFA)的支持，在云中实现最高性能的机器学习训练。4. G4 实例可达到 65 TFLOP 的 FP16 性能，成为小规模训练任务中无可争议的解决方案。5. Inf1 实例具有多达 16 个 Inferentia 芯片，这是由 AWS 专门设计和打造的高性能机器学习推理芯片。6. G5 实例配备多达 8 个 NVIDIA A10G Tensor Core GPU。7. Elastic Inference 允许客户将低成本 GPU 驱动型加速附加到 Amazon EC2 实例。以上计算资源可以覆盖所有的机器学习端到端的应用场景。其中 Trainium，Inferentia 都是近 4 年亚马逊自行研究的用于机器学习场景的专用芯片。

从生态链的集成融合看，在存储技术方面，截至 2022 年 12 月底，可以为 SageMaker 提供存储的服务有：1. S3 对象存储，提供行业领先的可扩展性、数据可用性、安全性和性能。各种规模和行业的客户可以为几乎任何使用案例存储和保护任意数量的数据，例如数据湖、云原生应用程序和移动应用程序。借助低成本的存储和易于使用的数据管理功能，可以降低客户的运营成本、提高数据的安全治理水平，同时满足特定业务、组织的和合规性要求。S3 可以作为存储层轻松地完成每秒数千次的事务处理，2. Amazon FSx for Lustre 提供完全托管式共享存储，兼具常用 Lustre 文件系统的可扩展性和性能。与 S3 集成可为共享文件系统带来高吞吐量以及一致的低延迟。3. Amazon Elastic File System (EFS)可随着添加和删除文件自动增大或收缩，无需管理或预置。EFS 能轻松地访问大型机器学习数据集或共享代码，无需预置存储或顾虑管理网络文件系统，同时 Amazon EFS 更易于使用且可扩展，并可提供机器学习（ML）和大数据分析工作负载所需的性能和一致性。EFS 为基于 Linux 的工作负载提供简单、可扩展的弹性文件系统，可与 AWS 云服务和本地资源配合使用。它可在不中断应用程序的情况下按需扩展到 PB 级，它旨在为数千个 Amazon EC2 实例提供大规模并行共享访问模式，可让客户的应用程序在一致、低延迟的状态下实现高水平的总吞吐量和 IOPS。Amazon EFS 是一项完全托管的服务，不需要对现有应用程序和工具进行更改，并通过标准的文件系统界面提供访问以实现无缝集成。Amazon EFS 提供标准和不频繁访问存储类。通过使用生命周期管理，可将 30 天未访问的文件自动移动到成本优化型不频繁访问存储类，让客户可以在同一文件系统中存储和访问经常使用和不经常访问的数据，而获得存储成本 85%的优化。Amazon EFS 也是一种用于在多个可用区 (Availability Zone)中存储数据以提供高可用性和持久性的区域性服务。4. Amazon Elastic Block Store (Amazon EBS)是一种易于使用且可扩展的高性能数据块存储服务，适合用于 Amazon Elastic Compute Cloud (Amazon EC2)。EBS 实现个位数毫秒级的延迟，以满足高性能存储需求。以上数据服务从性能，容量、数据吞吐、共享，完全覆盖了运营机器学习云平台所需的对数据的要求。

3.3.3 SageMaker 价格运营状况分析

在计算方面，SageMaker 以按需购买的方式帮助客户减小其总体拥有成本 TCO（Total Cost Ownership）[31], 同时省去客户的预置成本。通过 Savings Plan 服务，可以为客户承诺的未来一年或三年的算力消费金额提供高折扣（通常 60%以上）。最后对于成本极其敏感且可以容忍计算资源被临时抢占的客户提供竞价实例（Spot Instance），进一步大幅减少训练、推理的硬件费用（最高 90%）。除此之外，SageMaker 还提供了对指定机型，指定数量和时间段的免费试用服务。另外一个对客户成本友好的方面是，如果没有使用算力，也可以将机器置于关机状态，这段时间，客户只需要为对应机器的存储付费，实现了计算存储的灵活分离。以上所有计算服务的计价维度单位均为秒级，给客户提供了高精度的计价规则。

在存储方面，对于海量的数据，除了持久性，可靠性指标，其性价比是其技术运营能力的直接体现。两种最重要的数据存储类型分别是对象存储和文件存储。对象存储多是基于图片，视频的非结构化数据，文件存储则是偏向于文字的日志文件。对于面向海量存储的 S3，EFS 服务来说都有根据不同应用场景的不同定价模式。对于 S3 对象存储来说，可分为标准型，智能分层型，不频繁访问型，单可用区型，冷存储，超冷存储等共计 11 种存储类型，其价格也是随着访问性能逐渐降低。最高的 S3 标准型价格（0.1755 元/GB-月）是超冷存储（0.01082 元/GB-月）的 16.4 倍，除此之外对于对象存储还实施三个阶梯（前 50TB,50-500TB，大于 500TB）计价，用量越多单价越低，为客户提供了丰富的价格可选项。对于文件存储产品 EFS 而言，其提供了两种存储类别，标准型跟单可用区型，不同的存储类别都支持智能分类数据，将不频繁访问的数据分别放入价格更低的非频繁访问存储中。其标准高频访问的价格（2.359 元/GB-月）是单可用区价格（0.118 元/GB-月）的 19.9 倍。

3.3.4 SageMaker 品牌运营状况分析

企业品牌战略是企业运营的基本战略之一。传统品牌运营主要包含以下几个主要方面,即品牌传播、品牌定位、品牌包装、品牌文化、品牌忠诚度维护和

品牌创新。机器学习云平台作为高科技行业的典型代表，其品牌运营也有其特殊性。SageMaker 脱胎于亚马逊云科技，在其运营特点上跟亚马逊云的品牌策略有很多相似性。亚马逊机器学习云平台的品牌运营主要体现在针对重点客户群体实施以技术创新为主导的提高客户应用机器学习的能力建设上。SageMaker 的目标客户群体是有机器学习应用场景的企业客户。这些客户拥有实施机器学习的基础建模数据，但是没有相关的模型研发能力或大规模工程化能力。需要借助成熟的机器学习管道完成自身业务跟机器学习的关联，产生出具有商业智能的模型帮助企业提高决策效率。对于人群画像来说，SageMaker 的目标人群主要聚集在数据科学家，算法工程师，研发工程师和运维工程师等职位。SageMaker 在中国的品牌运主要围绕以下几个方面进行。

第一，打造持续创新的品牌形象。亚马逊云自 2011 年，每年十一月都会在美国拉斯维加斯举行声势浩大的新品发布会，并起名叫做 Re-Invent，从其命名就可以看出其主张创新的品牌形象。

第二，打造技术驱动创新的科技公司品牌。技术的核心是将先进的技术应用到真正的业务场景并创造价值，其技术实施者往往是个人或组织。对潜在的目标客户人群，SageMaker 通过培训，认证机制为技术人才的技能能力背书。通过部分功能的免费试用降低技术人员在学习新技术的门槛跟成本。

第三，与其他行业的标杆公司联合打造具有行业渗透性的品牌影响力。例如 SageMaker 在通用公司的应用，帮助 SageMaker 在能源行业积攒了口碑。SageMaker 在联想集团的试用体验帮助其扩大了在制造业客户的影响力。这种通过与生态链合作伙伴的联合，共同打造具有行业属性知名度的策略已经覆盖了包括能源，制造，金融，自动驾驶，广告等行业。

3.3.5 SageMaker 生态链运营状况分析

SageMaker 对生态链合作伙伴的运营主要体现在以下几个方面：1. 对生态链合作伙伴进行管理和激励，包括生态链合作伙伴招募、资源分配、奖励政策等。2. 鼓励生态伙伴创新和协同。3. 提升生态伙伴的积极性和参与度。在 2022 年第四季度，亚马逊云科技和微软都推出了面向人工智能的生态链合作伙伴计

划，认识到与生态链合作伙伴合作推动客户使用其人工智能产品的重要性。而就 SageMaker 而言，其生态链有以下两种形式。

第一种生态链形式是 SageMaker 与云平台其他云服务的生态关系。除了托管的机器学习云平台 SageMaker，亚马逊云科技对于不同机器学习场景均有覆盖。从面向底层的只提供基础设施服务，到中间部分的面向 API 的服务，再到最上层的全 SaaS 服务，SageMaker 以不同的颗粒度满足着各个维度的场景化需求，其通用解决方案架构如下图 3.9 所示。

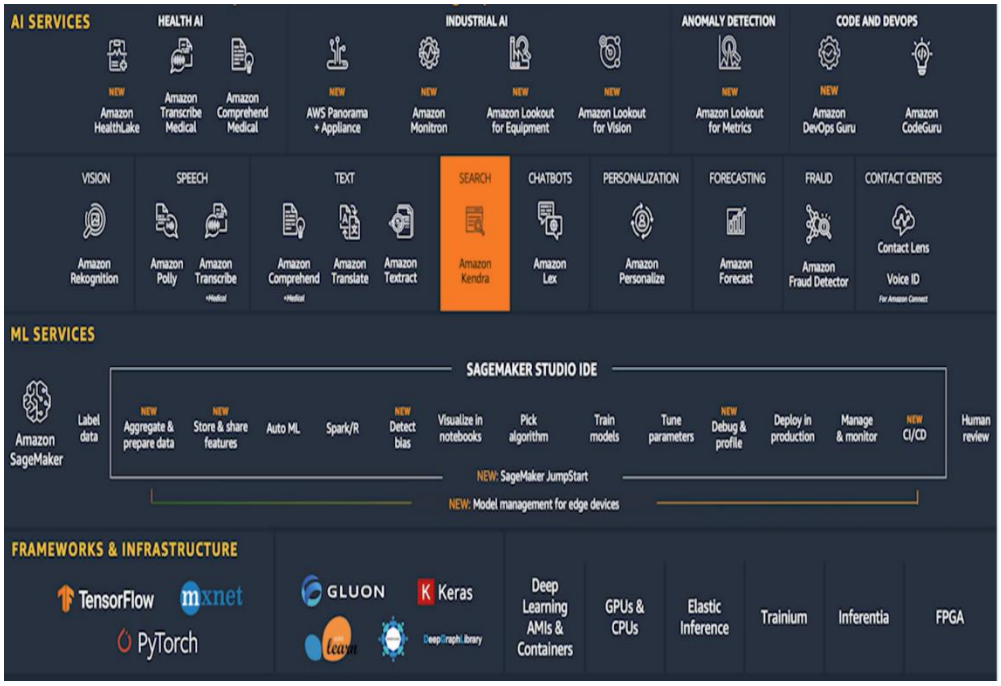


图 3.9 亚马逊云计算机器学习解决方案概览-源自 SageMaker 官网

SageMaker 作为一个全托管的机器学习云平台，植根于亚马逊云平台的生态链系统。其定位于生态链的顶端，而运营整个生态链，使其互相赋能，尤其是高效整合底层生态服务特别考验云平台的综合运营能力。该生态链的构建通过融合云平台的其他服务，形成了完整的产业链和价值链。这个生态系统涉及到必须的基础服务如，计算，存储，网络，安全等服务。如在机器学习的场景中数据通过网络接入云平台，面对不同的网络要求选择的服务也不尽相同，如对于网络延迟和稳定性有高要求的金融保险行业，专线服务（Direct Connect）是首选，对于大量的物联网设备的数据上传，IoT Core 服务几乎是必选项，对于跨区域的加密公共互联网通信来说 Site to Site VPN 是最佳选择。对于数据安

全来说，也是企业级客户最为关注的头等大事，什么角色可以访问什么数据需要用到 IAM 服务，对数据的加密需要用到 KMS 服务，对于网站的证书需要 ACM 服务，而权限系统里普遍使用的密钥则需要使用 Secretes Manager 服务。除了上文提到基础服务，还有数据库，数据分析等增值服务，互相搭配才能构成整体的解决方案版图，满足客户的细分业务需求。

第二种生态链形式是 SageMaker 应用商城的生态链。应用商场是一个云上的在线商店，有偿提供适用于 SageMaker 的机器学习算法、模型和其他相关工具。SageMaker 应用商城提供预构建的机器学习模型、算法和其他工具，可以轻松集成到客户的机器学习流程中。市场上的模型和算法由亚马逊和第三方开发人员开发，并经过性能测试和数据验证以确保与 SageMaker 兼容。使用 SageMaker Marketplace 可以轻松浏览和搜索所需的模型和算法，并将其直接部署到隔离的 SageMaker 环境中。应用市场提供了各种模型和算法，用于图像分类、自然语言处理、目标检测等任务。SageMaker Marketplace 也为数据科学家和开发人员提供了一个平台，可以销售自己预构建的机器学习模型和算法。这不仅可以为贡献者带来经济上的收入，也可以使他的工作得到更广泛的传播和认可。总体而言，SageMaker Marketplace 提供了一个方便高效的方式来访问和部署预构建的机器学习模型和算法，加速机器学习应用程序的开发。

3.3.6 高层调研整理汇总

为了进行多维度，深层次的分析，从而更加客观的反映 SageMaker 在中国运营的现状。作者对亚马逊云科技的多位高层领导进行了访谈。访谈针对 SageMaker 在中国运营的关键领导岗位，即亚马逊云科技中国执行董事、产品总监、生态部总经理。作者针对不同的高管背景设计了不同的采访提纲。对中国执行董事，作者以当前国内云计算的发展现状为访谈切入点，逐步发散并深入到亚马逊云科技在技术，成本运营的探讨。对于 SageMaker 中国区的产品总监，讨论聚焦于当下被热议的生成式 AI 的产品覆盖，以及亚马逊云科技品牌运营的策略上。针对生态部总经理的访谈则以跨国知名外企在海外生态运营的成功经验开题，深入探讨了国内生态建设的现状以及基于行业生态拓展的运营策略。具体访谈提纲请参见附录 1，以下是针对每位高管的调研总结。

执行董事将亚马逊云科技在中国运营的策略从技术、价格、品牌和生态链四个维度进行了客观评价。在中国如火如荼的数字经济浪潮中，云计算凭借着提供“拎包入住”的创新生态链环境、“按需付费”的高灵活性和成本优势以及降低新技术使用门槛这三大优势，正成为数字经济的“技术底座”，扮演着越来越重要的角色。所以对亚马逊云科技来说，技术运营尤其重要的，这也是亚马逊云科技的中国领导团队一直在积极推进新服务落地中国的首要原因。由于云计算的规模效应结合用户使用过程中的“按需付费”，可以有效降低客户 IT 资源的整体拥有成本。但是在中国的实际运营 SageMaker 过程中，国内友商可以通过非常灵活的定价提供有竞争力的成本优势。在品牌方面，虽然亚马逊云科技在全球范围内有较好口碑和市场占有率，但是国内的云厂商在各自的领域有很高影响力。最后，亚马逊云科技非常重视生态链建设，这也是为什么 AWS 在全球拥有数万家生态链合作伙伴，从而形成了一个横跨各领域的复杂生态合作伙伴网络简称 APN（AWS Partner Network）。但是在中国运营过程中，由于进出口的有关限制，我们在生态完整性方面仍在持续发力。

SageMaker 中国产品总监对生成式 AI 的火爆出圈并不感到意外，但是随着模型的越来越大，降低其训练成本显得至关重要，因此从技术运营层面通过并行的方式进行分布式训练跟推理就变得非常必要。从支持机器学习模型训练跟推理的角度说，拥有更强大算力的基础设施又是生成式 AI 的生态保障。亚马逊云科技研发了极具性价比的专属训练和推理芯片，目前正在积极的向国内引进。同时国内的友商也在 2017 年陆续投入到专属芯片的研发中，并有各自的云产品已经完成上线部署，进一步增加了对亚马逊云科技在中国运营的挑战。

亚马逊云科技中国生态部总经理在访谈中表示，截至 2022 年底，只有 5%到 15%的 IT 投入应用到云端，可见这一细分市场无疑拥有非常巨大的发展前景，伴随着大量的传统企业上云，一个强有力的生态系统布局就变得不可或缺。亚马逊云科技依赖其在海外的巨大影响力通过与众多知名企业联合开展合作，加速生态链合作伙伴在中国的迅速成长。但是在某些垂直领域，如金融、医疗、教育的渗透率方面仍然有较大发力空间。

3.3.7 基于 SWOT 分析法的 SageMaker 在中国运营策略

SWOT 分析法由 S 优势（Strengths）、W 劣势（Weaknesses）、O 机会（Opportunities）、T 威胁（Threats）四要素构成。用于评估企业、组织、项目、个人的优势劣势，通过将机会跟优势最大化、劣势与威胁最小化，为制定企业战略或个人发展提供指导。基于以上对 SageMaker 在中国从技术、价格、品牌、生态链运营状况的分析，将 SageMaker 的 SWOT 分析概括如下：

优势方面（Strengths）：亚马逊云是业界公认的第一梯队云服务提供商，截至 2022 年底，亚马逊云计算占有世界 40%的 IaaS 市场份额。除了基础设施方面，在大数据、网络安全、应用开发、机器学习，数据分析领域覆盖两百多个服务，每年开发迭代的新功能超过五千个。与其合作的生态链合作伙伴数以十万计。从其官方的技术博客可以看到拥有众多世界 500 强企业的成功故事覆盖不同行业，如通用，丰田，阿迪达斯，耐克等。以上庞大的基础，为机器学习场景的应用提供了先决条件，也为亚马逊机器学习云平台的持续迭代更新提供了足够的场景化需求。

劣势方面（Weaknesses）：亚马逊云起初植根于自己的电商平台，以两个披萨的开发文化迅速占领了云计算市场。但是对于企业级公司的办公软件以及 IT 系统，不如微软的影响力。对于开源贡献来说不及谷歌。另外，亚马逊云进入中国较晚，在中国企业的应用基础较弱。对 SageMaker 而言，其中国跟海外的服务存在一定差异，导致场景覆盖存在一定短板，需要客户的定制开发才能满足客户需求。由于亚马逊云提供的云计算服务较多，彼此间可以互相独立或组合使用，导致客户的学习曲线较高。最后，中国有比较严格的票据制度，内容审核机制，以及中国客户特有的技术需求，这些细分领域的特殊需求比较难被以通用解决方案见长的云厂商所重视。

机会方面(Opportunities)：目前公有云市场尚未饱和，中国大部分互联网公司基本实现了云化，但是还有相当一批传统企业还在搭建自己的数据中心，这些客户在国家推进云计算的趋势下也在逐渐转型。2022 年生成式 AI 技术席卷全球，越来越多的商业机构将目光投向了 AI 产业，而 AI 产业天生与公有云平台紧密依赖，未来场景值得期待。最后，数字转型的热潮还在涌动，所有企业

都在试图通过自身数据了解自身问题和发展方向。利用公有云服务完成数字转型并发现隐藏在数字背后的商业洞察是业界普遍认可的提高企业自身竞争力的有效手段。

威胁方面(Threats):其他国际巨头在中国的发展以及本土厂商的迅速崛起给 SageMaker 在中国的运营提出了巨大挑战。国内厂商在跟进国内政策方面的速度极快。依赖中国健全的产业链规模，国内云厂商的服务定价更加友好。国内云厂商在跟国内客户的商业优势互换、合作上也更容易找到合作共赢的机会。

结合 PEST 宏观分析，SageMaker 在中国的运营策略概括为如下图 3.5 所示的波士顿矩阵。

外部因素 / 内部因素	机会 (O)	威胁 (T)
	• 生成式AI 热度驱使人工智能云平台的应用增多 • 仍有大量企业还没有完成云化 • 数字化转型仍在进行中	• 来自本土厂商竞争日趋激烈 • 客户定制化服务增多 • 国内厂商更容在商业互换中找到合作机会 • 国内厂商价格定义更灵活
优势 (S) • 国际区SageMaker技术领先，市占率高 • 产品覆盖范围广，方案灵活 • 国际区生态丰富，资源充沛 • 中国区客服质量高	OS 积极进攻策略 • 开发国内客户的海外需求 • 将国外的成功数字化转型经验带到国内 • 将国外知名企业复杂的上云过程方案化，推广到国内	TS 差异化策略 • 利用海外SageMaker的强大生态满足国内客户的部分需求 • 通过多个服务的长期承诺用量争取折扣，降低客户整体成本 • 寻找客户出海的可能性合作，减小客户对价格的敏感度
劣势 (W) • 价格较贵 • 客制化功能开发较弱 • 受进出口限制，GPU硬件资源不足 • 缺乏中国明星客户背书 • 本土化生态建设不足	OW 弱点强化策略 • 通过成功方案的平顺实施，减少客户对成本的担心 • 通过自研的芯片替代受限制的GPU资源 • 通过跨国公司在中国落地其国外SageMaker经验做成功故事突破 • 联合国内垂直行业厂商做第三方方案的客制化开发	TW 防守/关停策略 • 通过长期预留实例给客户最低价格 • 通过客户在国内厂商的失败体验，赢得机会 • 对初创公司给予重视，做前期投入 • 通过对新政策的已知解决方案发掘，吸引新客户

图 3.5 结合 PEST 宏观分析的波士顿矩阵

从以上矩阵分布可以看出，在外部整体环境因素利好，自身海外优势明显的现状下，SageMaker 在中国采取了 OS 增长策略。但是随着本土厂商在价格，生态链竞争日趋激烈的背景下，需要对 SageMaker 在中国的运营策略进行优化。而其优化的方向为 OW 策略，即弱点强化的运营策略。OW 运营策略需要企业苦练内功，抓住机会弥补短板、强化弱点，提升竞争力。因此需要找到 SageMaker 在中国运营的问题并加以解决。

4 SageMaker 在中国运营策略存在的问题分析

通过对高管的访谈，以及 SWOT 的综合分析，作者对亚马逊云平台在中国运营的核心问题归纳为以下四个方面。第一，缺失关键技术功能，导致无法完成平台闭环的技术运营问题。第二，对客户总体拥有成本不友好的价格运营问题。第三，无法大量复制成功案例，做到快速一到一百市场传播的品牌运营问题；第四，生态链发展滞后、存在明显短板的生态链运营问题。接下来分别从技术，价格，品牌、生态链角度深入分析其形成原因。

4.1 技术运营中的不足

4.1.1 国内外功能发布不同步

SageMaker 在国内的短板主要体现在两个方面，一个是平台自身功能缺失，另一个是生态链依赖的其他服务不足。

就计算服务来说，截至 2022 年 12 月底，中国区虽然上线了亚马逊云科技专为推理任务设计的加速芯片 Inf1，但并没有上线当下主流的 Nvida A100 Tensor Core GPUs P4 系列机型(由于 2022 年 9 月美国商务部发布的芯片限制出口法案)和 NVIDIA A10G Tensor Core GPU 的 G5 系列机型。相对于最新一代的 GPU 计算实例，目前中国服役的训练芯片不仅在张量计算的吞吐率，还是性能上都有较大差距。另外，专为深度学习设计的 Trainium 芯片在中国也没有上线。对于其他架构的计算资源如 CPU，ARM 在中国还停留在第六代（C6，R6），而其海外已经普遍上线第七代（C7，R7）。从 SageMaker 产品的生命周期迭代来看，其功能也有比较明显的滞后。自动标注功能（SageMaker ground truth）可以帮助客户完成自动的数据标注工作，减小数据处理的时间跟成本；托管的强化学习（Reinforcement learning）可以帮助在非监督的训练场景下减少环境的配置工作；弹性推理（Elastic inference）可以加快吞吐量提高推理的速率。以上提及的三个服务目前在中国区均没有上线。

除了上文列出了三个缺失服务，SageMaker 在中国仍有十四个功能缺失。如 Studio Lab，JumpStart，distributed training libraries 等，还有两个功能使用上

有限制，如 Marketplace China。SageMaker 在海外的 28 个区域中自 2021.1.1 日至 2022.12.29 总共完成了 79 个新功能的发布，而同期在中国的发布只有 5 个。考虑到海外的功能发布公告是按照区域进行划分的，而中国则是合并发布的，所以 2022 年针对中国的发布有 10 个，分别来自北京区跟宁夏区域。从发布速率来看海外有明显优势。针对同样的功能发布，国内也和海外有时间的延迟。

表 4.1 2021.1.1-2022.12.30 SageMaker 新功能及其发布时间

-数据源自 SageMaker 新功能官网

功能名称	中国区发布时间	国际区发布时间	延期天数	年度平均延迟
Amazon SageMaker Studio and SageMaker Notebook Instance now come with JupyterLab 3 notebooks	2022/10/10	06/06/22	-126	
SageMaker Studio now supports Glue Interactive Sessions	9/14/22	9/13/22	-1	
Now track user identity for API calls from Amazon SageMaker Studio in Amazon CloudTrail	9/7/22	9/8/22	1	-46.8
Amazon S3 increases the maximum number of S3 Access Points and adds support for Amazon Redshift, Amazon CloudFront, and Amazon SageMaker Feature Store	8/24/22	7/27/22	-28	
Now prepare data and build models using TensorFlow 2.6 and PyTorch 1.8 in Amazon SageMaker Studio Notebooks	2/11/22	11/23/21	-80	
Amazon SageMaker now supports inference testing with custom domains and headers from SageMaker Studio	11/18/21	11/4/21	-14	
SageMaker Studio enables interactive Spark based data processing from Studio Notebooks	10/18/22	9/21/22	-27	
Now Use Lifecycle Configurations to Customize Amazon SageMaker Studio	9/24/21	9/24/21	0	
Amazon SageMaker now supports inference endpoint testing from SageMaker Studio	9/23/21	9/24/21	1	-62.75
Amazon ml.Inf1 instances are now available on Amazon SageMaker in the Amazon Web Services	7/8/21	4/20/20	-444	
Now Use Tags to Track and Allocate Amazon SageMaker Studio Notebooks Costs	4/14/21	4/15/21	1	
Now create Amazon SageMaker Studio presigned URL with custom expiration time	2/22/21	2/20/21	-2	
Now launch Amazon SageMaker Studio Notebooks backed by Spark in Amazon EMR	1/7/21	12/21/20	-17	

从上表 4.1SageMaker 的功能发布情况来看，新服务在中国明显有发布延迟。从年度的统计维度来看，SageMaker 在 2022 年的延迟天数上有明显改善，提高了近 25.8%，表现出了积极追赶的态势,但是仍用 46 天的推迟。从具体的功能来看，可以分成三大类。第一类，SageMaker 与硬件资源的集成，如 SageMaker 适配推理芯片 ml.Inf1 的集成比海外晚了 444 天。第二类：SageMaker 与软件系统集成，如 SageMaker 与 JupyterLab3 的集成延迟了 126 天，与主流机器学习框架 PyTorch1.8 的集成拖延 80 天，与大数据框架 Spark 的集成晚了 27 天。第三类：SageMaker 与其他生态链服务的集成，如 SageMaker 与 S3 存储服务的集成有 28 天的延误。对于中国没有发布的功能

如：2022.7.15 发布于海外的“Amazon SageMaker 新增用于模型部署的 ml.g5、ml.p4d 和 ml.c6i 实例”，是因为其依赖的 P4d, ml.g5 计算服务没有在中国落地而被迫推迟的。经过以上分析，不难看出，作为云平台集数据与算力之大成者的机器学习云平台 SageMaker，其功能复杂，依赖程度高，尤其当涉及到跨技术栈与硬件，软件系统和生态链集成时，对其在中国的运营能力提出了极大挑战。

4.1.2 对于中国元素场景覆盖不足

中国作为世界第二大经济体，其世界影响力与日俱增，而针对中文的机器学习云服务也是机器学习云平台的一个重要组成部分。其中比较常见的如中文的自然语言理解，中文的文字识别，中文的翻译，以及文字跟语音的互相转换。而这些服务目前在中国的落地情况并不理想。如 Translate（文字翻译）, Comprehend（文字理解），Textract（文字提取）并没有上线。而已经上线的 Polly，Personalize，Transcribe 对中国场景的支持程度也不令人满意，如 Personalize 目前不支持对非结构化的中文进行推荐。Transcribe 还不能自动识别语言种类，Polly 无法在北京区使用。由此可以看出在托管的机器学习服务中，还是欠缺的。亚马逊云科技的所有创新均来自于客户，并把“客户至尚”写入公司的领导力准则，所以其 SageMaker 的客户故事可以反映出其产品的战略方向。在 SageMaker 海外官网的客户故事中，如下图 4.10 所示，只看到了一个中国元素公司-Lenovo（联想），而其宣传的服务（SageMaker Edge Manager）中国还没有落地。另外，在 SageMaker 的中国官网，没有看到任何用户故事。足见不论从产品的中文本地化还是头部用户的成功案例，都显示出覆盖性不足的短板。

图 4.10 SageMaker 国际区大客户代表-截图源自 SageMaker 客户故事官网

4.1.3 底层设计灵活，有较大学习曲线

SageMaker 机器学习平台诞生于 2017 年 11 月 29 日，截至 2022 年 12 月 30 日，五年间累积发布了 280 个新功能，如下表 4.2 所示。从功能发布数量以及年度的关联度看，自 2021 年, SageMaker 的发布迭代速度明显加快，几乎以成倍的速度在发展，这直接导致其学习成本越来越高。

表 4.2 SageMaker 2017-2022 年发布功能数量统计-数据源自 SageMaker 新服务列表

年度	SageMaker 新功能数
2022	79
2021	74
2020	38
2019	41
2018	46
2017	2
总计	280

除了功能数量，能直接反映一个产品的复杂度以及学习曲线的数据就是其使用说明文档的数量以及篇幅长短。SageMaker 面向客户的文档有四类，每一

类都有十一种语言支持。第一种文档，面向开发人员的使用说明书（英文版 3870 页）；第二种，API 接口说明文档（英文版 1914 页）；第三种，面向 Python 开发语言的 SDK 说明书；第四类，面向 Boto3 的 SDK。以上四类文档一方面可以有效说明 SageMaker 的归档工作非常细致，另一方面也反映了其设计灵活性以及较高的学习门槛。同时叠加在中国的新功能上线较慢，导致在功能需求管理上，对应的文档版本上都提出了很大运营挑战，所以用户在使用中国 SageMaker 产品时会感到有一定上手难度。

4.2 价格运营中的不足

从亚马逊云科技对 SageMaker TCO 使用成本的研究报告中可以看到，如表 4.3 所示，以三年的周期作为分析维度，通过 SageMaker 构建、训练、模型部署的全生命周期中，平均基础设施的花费占到了总体拥有成本的 92.18%，其中存储、计算占到了基础设施花费的 95%以上，所以对于 SageMaker 的价格运营能力，主要体现在对计算，存储的综合价格运营能力上。

3 year TCO Analysis					
X-Large Scenario	Amazon SageMaker	Amazon EC2	Amazon EKS	Amazon SageMaker TCO v/s EC2	Amazon SageMaker TCO v/s EKS
TOTAL COSTS (Build + Train + Deploy)	**$2,460,412**	**$10,594,181**	**$5,342,908**	**-77%**	**-54%**
Total Build Costs	**$490,786**	**$1,374,026**	**$1,374,026**	**-64%**	**-64%**
Infrastructure Costs	$426,710	$311,049	$311,049		
Operational Costs	$64,076	$640,755	$640,755		
Security & Compliance Costs	$ -	$422,222	$422,222		
Total Train Costs	**$298,770**	**$3,704,034**	**$1,751,164**	**-92%**	**-83%**
Infrastructure Costs	$234,695	$2,641,057	$688,186		
Operational Costs	$64,076	$640,755	$640,755		
Security & Compliance Costs	$ -	$422,222	$422,222		
Total Deploy Costs	**$1,670,856**	**$5,516,120**	**$2,217,718**	**-70%**	**-25%**
Infrastructure Costs	$1,606,781	$4,453,143	$1,154,740		
Operational Costs	$64,076	$640,755	$640,755		
Security & Compliance Costs	$ -	$422,222	$422,222		

表 4.3 TCO of SageMaker–源自 AWS SageMaker 海外官网

4.2.1 高性价比计算资源和成本优化服务缺失

有更高性价比的计算资源如 Nvida A100 Tensor Core GPUs 的 P4 系列计算

类型(由于美国芯片制裁法案)和 NVIDIA A10 GPU 的 G5 系列机型目前还没有在中国上线。专门为深度学习设计的 Trainium 芯片在中国也没有时间表。其他架构的计算资源如 CPU，ARM 中国还只能用第六代机型（C6，R6），而海外已经普遍上线第七代（C7，R7）性价比更高的产品。虽然这些差距会随着时间的推移慢慢补齐，但就目前而言，短板效应尚且明显。除了性价比较高的新计算资源，目前国内还没有面向所有计算资源打折的 Savings Plan 服务。该服务将客户承诺的一个长期的资源使用量转换为较低的整体计算价格折扣。另外 Savings Plans 支持在 Compute、EC2、Fargate 和 Lambda 上的联合购买计划，并提供一种更灵活的购买方式，因为它可以与现有的 Reserved Instance（RI）进行组合使用，因此可以获得更好的价格优惠。Savings Plan 的缺失对计算资源价格敏感的用户非常不友好。

4.2.2 免费试用服务以及服务优惠存在缩水现象

第三方独立咨询机构-字母点评[32]通过对 20 个应用过 SageMaker 不同行业的企业进行了调研，得出的整体能力评级如下图 4.12 所示。

图 4.12 字母点评对 SageMaker 综合评级

价格因素明显拉低了对 SageMaker 的整体评价。虽然 SageMaker 完全按照按需付费的计价模式运营，仍然有 33%的用户认为价格较高，如下图 4.13。

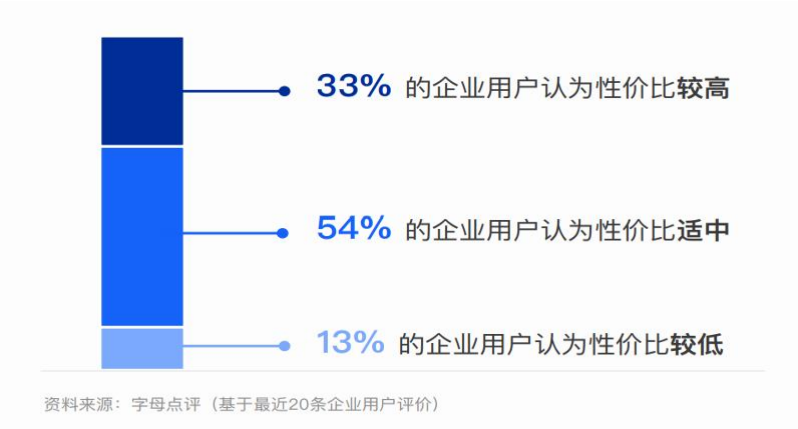


图 4.13 企业客户调查中对 SageMaker 的价格评价

正如 SageMaker 总体拥有成本研究报告中提到的，SageMaker 有两部分较大开支，一个是 SageMaker 自身计算、存储的托管成本，另一个是其生态链的其他服务成本，如 SageMaker Ground Truth 等。在国内，目前只有 SageMaker 提供免费试用功能，而在海外除了 SageMaker, SageMaker Ground Truth 也提供免费试用。然而就 SageMaker 中国的免费试用而言，相较于其海外也有一定程度的缩水，如图 4.14 和 4.15 所示。对于计算资源，国内只提供 t2 的上一代机型作为免费笔记本实例，而海外采用的是更新一代的 t3 类型（t3 较 t2 有 30%性价比优势）。海外的免费试用还包括 25 小时 ml.m5.4xlarge 计算资源做机器学习前的数据处理。在存储方面，海外的 SageMaker 可以让客户存放 1000 万写入单位，1000 万读取单位，容量为 25GB 的特征数据。以上免费试用功能，虽然不会对客户的成本有较大影响，但是可以看出，SageMaker 在免费试用环节，国内跟国外的策略存在明显差异。

图 4.14 中国区 SageMaker 免费试用条目

图 4.15 国际区 SageMaker 免费试用条目

对于生态链的其他服务成本来说，海外有 80 个产品可供免费试用，而中国只有 40 个，如下图 4.16。

图 4.16 国内，海外免费套餐截图

在海外直接降价或由于技术升级导致对客户使用成本有间接帮助的服务数量统计如下图 4.17 所示：

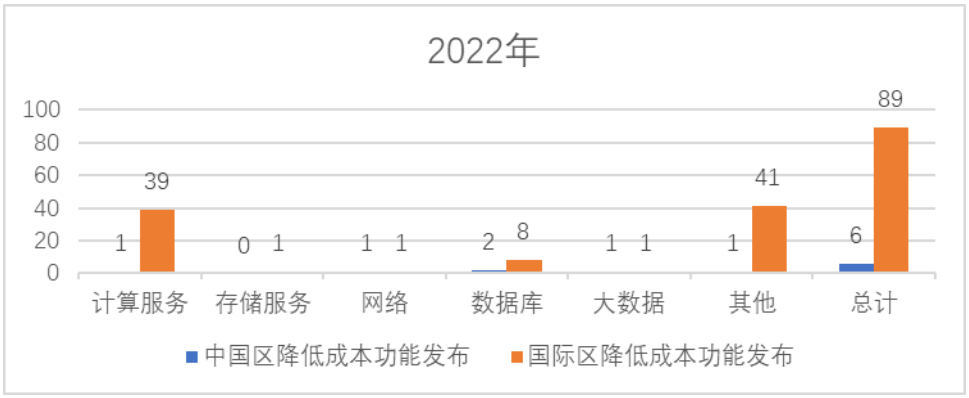


图 4.17 2022 年中国与海外降价服务发布对比

2022 年在中国关于定价下调的服务发布有 6 个，而海外累计达到 89 个。直接降价的服务中国有 1 个，而海外有 5 个，统计数字如下表 4.3 所示。对于成本的主动降价海外有明显的运营优势。

2022年	计算服务	存储服务	网络	数据库	大数据	其他	总计
中国区直接降价的服务	0	0	1	0	0	0	1
国际区区直接降价的服务	3	0	1	0	0	1	5

表 4.3 2022 年国内与海外直接对服务降价的发布次数对比

就降价幅度而言，只有“PrivateLink，Transit Gateway 在同一个可用区的免费网传输”跟海外同步发布。其他只在海外发布的，可以间接降低客户使用 SageMaker 成本的服务平均降幅高达 58%。举例来说，假设 SageMaker 的训练数据由 IoT 设备采集并通过上一代的通用性计算实例 C6g 做模型训练，那么 IoT 设备管理功能降价 80%，同时叠加 C7g 的性价比相较于 C6g 的提升 40%，利用 SageMaker 机器学习平台的总体花费会有比较可观的价格优化。而这些降幅对国内用户并不可见，也就没有形成价格的运营优势。下表 4.4 是 SageMaker 的降价服务以及发布时间。

	降价详细内容	发布时间	降幅
国际区区直接降价的服务	AWS Config announces a price reduction up to 58% for conformance packs.	14-Sep-22	58%
	Amazon EC2 C7g instances powered by AWS Graviton3 processors now available in US East (Ohio) and Europe (Ireland)	10-Mar-22	40%
	AWS IoT Device Management announces an 80% price reduction for Secure Tunneling	7-Jun-22	80%
	Price reductions on Amazon EC2 instances running SUSE Linux Enterprise Server (SLES) OS	1-Jun-22	52%
	AWS Announces Data Transfer Price Reduction for AWS PrivateLink, AWS Transit Gateway, and AWS Client VPN services	7-Apr-22	100%
中国区直接降价的服务	Amazon Web Services announces data transfer price reduction for Amazon PrivateLink, and Amazon Transit Gateway services	7-Apr-22	100%

表 4.4 2022 国内，海外直接降价的云服务及其发布时间

4.3 品牌运营中的不足

SageMaker 在国内的品牌运营跟亚马逊云科技共享同样的品牌运营策略。其主要原因是体现全球的统一标准，让全球的客户可以无差异的享用亚马逊云所带来的一致体验。另一个原因是云平台属于生态经济，SageMaker 是云平台的集大成者，充分利用亚马逊云的市场领导力可以帮助 SageMaker 获得品牌背书。为了更好的适应中国机器学习的落地场景，亚马逊云科技在 2018 年于上海成立了上海人工智能研究院。为机器学习在中国的落地持续发力，同时加强机器学习云平台的市场影响力。亚马逊云科技每年都会在重点城市开展技术峰会。针对每年十一月美国拉斯维加斯 Re-invent 大会，国内也会组织高规格的线

下峰会跟进，为新发服务做市场的技术宣传。亚马逊云科技拥有自己的官网，微信公众号和完备的视频分发网络帮助其在媒资渠道持续扩大品牌影响力。借助以上市场活动，亚马逊云科技可以收集完整用户画像，为商务的下一步扩展提供商机。亚马逊云科技为加强品牌运营能力，设置了机器学习售前解决方案架构师，机器学习产品经理，以及机器学习应用科学家等技术职位以加强其在中国市场的品牌影响力。SageMaker 在中国对新用户实施指定机型的免费试用功能。完备的技术文档和机器学习方法论给入门用户提供完备的上手指南。上手实验的内容也会定期更新满足新场景的覆盖。虽然亚马逊云科技在机器学习云平台处于国际领先地位，但是在中国市场的影响力相对于华为云，阿里云，腾讯云还有差距，尤其是在本土金融，教育，医疗等行业。具体体现在缺少标杆企业的灯塔效应上。

4.3.1 标杆企业成功案例的博客曝光量低

技术博客的曝光是提高示范效应的有利手段，从 SageMaker 技术博客的运营角度看，曝光比例越高其市场知名度越高。作者统计了自 2018 年亚马逊云科技国内的官方博客，并绘制了统计表，如表 4.4 所示。从统计表的数据看，SageMaker 的曝光率自 2019 年之后都远高于其它服务。而随着疫情自 2020 年的爆发，风控对线下举办技术类交流影响较大，博客成为技术运营的有力支点。SageMaker 转而重点通过技术博客输出其技术领先的品牌影响力。但是仔细分析 SageMaker 的技术博客，并没有找到一篇是中国客户使用国内 SageMaker 的成功故事。从亚马逊云科技的客户案例精选中也只找到了一家中国公司（大宇无限）使用 SageMaker 的案例。案例中有如下描述“考虑到以上 IT 挑战和需求，大宇无限选择了全球卓越的云计算服务商亚马逊云科技作为自己的出海搭档。大宇无限充分利用亚马逊云科技机器学习平台 Amazon SageMaker 进行精准视频推荐”。从以上内容可以看出，其宣传对象也是海外的 SageMaker 服务，也间接印证了技术博客缺少中国明星企业代言的判断。与之形成鲜明对比的是海外 138 个 SageMaker 客户精选案例，不乏 BP（英国石油），通用，Expedia，汤森路透，韩国现代等世界顶流公司。

年份	所有云服务新功能	SageMaker新功能	SageMaker博客曝光	总博客曝光量	SageMaker曝光比	SageMaker新功能占比
2022	1936	79	26	426	6.10%	4.08%
2021	2090	74	75	692	10.84%	3.54%
2020	2147	38	14	554	2.53%	1.77%
2019	1817	41	15	346	4.34%	2.26%
2018	1369	46	1	248	0.40%	3.36%

表 4.4 统计信息来自于亚马逊官网 https://aws.amazon.com/cn/blogs/china

4.3.2 标杆企业成功案例在顶级会议曝光量低

从标杆企业在顶级会议曝光度看，SageMaker 的有影响力的曝光也相对较少。SageMaker 第二重要的品牌运营手段是每年的顶级峰会。2022 年亚马逊云科技在中国共组织了四场旗舰会议。分别是 2022 Re：invent 全球大会，2022 Summit 中国峰会，2022 创新大会-人工智能新引擎，2022 创新大会-云基础架构。2022 Re:Invent 的发布中，关于 SageMaker 的分会场有 10 个，客户有三星电子，汤森路透，Zalando，NatWest，并没有中国公司。在 2022 Summit 中国峰会中，SageMaker 被分在人工智能创新引擎板块，涉及的分会场有三个，涉及的唯一的中国公司分享是 OPPO 的出海业务。在 2022 创新大会-人工智能新引擎中，第五分会场“大规模机器学习实战”中有两个关于 SageMaker 的技术分享但是均没有客户帮助站台宣传。2022 创新大会-云基础架构没有涉及 SageMaker 的话题。从以上四个中国旗舰线下峰会的市场活动看，中国客户使用国内 SageMaker 的案例十分有限。

亚马逊云科技中国国际客户及生态链合作伙伴，生态系统事业部总经理沈涛在 2021 年 11 月接受《哈佛商业评论》副主编钮键军采访时说，亚马逊云科技在中国实行“三驾马车”的战略。首先是中国客户深耕本地，第二是海外客户植根中国，第三是中国客户的成功出海。所以对 SageMaker 的运营也都应该围绕这三个场景展开。但基于以上技术博客和顶级会议对 SageMaker 的曝光度看，其深耕的中国本土客户的数量还是相对较少，更是缺少明星标杆企业的示范效应。

4.4 生态链运营中的不足

越是复杂的集成性系统越需要生态链的运营。生态链运营的特点充分体现

了木桶效应-即上层产品的成功不取决于其单一功能的卓越，而是对于普遍用户而言的一般功能是否存在致命短板。从生态链运营的视角看，负责 SageMaker 生态建设的部门需要持续不断的审视和评估生态链系统的短板，从供应链的上中下游，到集成商的硬件软件，甚至从局部到全球。机器学习云平台是万亿级别体量的市场，涉及的领域广，行业多，场景复杂。企业必须使用有效的决策模型、风险评估模型和风险优化模型，保障生态链运营的有效性。这将帮助决策者确定最优的投资策略，以使失败的可能降到最小。如何协调上下游利益，共同在机器学习云平台上发挥最大的客户价值是云厂商需要迫切解决的问题。亚马逊云科技在国际生态链运营方面非常成功，在中国也在持续发力中。

虽然国内有面向 API 开箱即用的 SaaS 服务如 Polly（将文字转换为逼真语音的服务），Transcribe（语音转文字），Personalize（精选推荐服务），但中国客户最经常使用的功能如 Translate(文字翻译), Comprehend（文字理解），Textract(文字提取)目前没有上线。这也反映出 SageMaker 在中国元素的支持方面有待进一步提高。Marketplace 的整体生态建设也进展缓慢，最突出的问题是没有针对 SageMaker 的应用市场生态。

4.4.1 本土化生态链建设不足

SageMaker 是全托管的机器学习云平台，其生态链运营重点聚焦在如下两个方面。一个是 SageMaker 跟自身云平台除了计算，存储的其他相关服务的生态链建设，如网络，安全，大数据服务等；另一个是与亚马逊云科技集成的生态链合作伙伴的生态链建设。目前这两个方面在国内发展都相对缓慢。

SageMaker 与亚马逊云科技自身生态链方面的短板主要体现在：第一，部分 SaaS 的 AI 服务没有落地中国。第二，部分非计算，存储的服务没有落地，导致整体的方案效果不佳。针对第一点，往往是 SageMaker 的商机，没有 SaaS 服务，可以利用 SageMaker 来做定制化的开发，但是违背了 SaaS 开箱即用，灵活付费的初衷，客户面临两难抉择。针对第二点，比如 Savings Plan 服务可以在海外实施基于整体花费承诺的跨机器家族的优惠，客户不用手动去操作预留实例的转换来节约成本，但是该服务在国内还没有上线，这对客户的运维提出了挑战。而这个服务的缺失，并没有第三方应用商店的方案可以绕道解决。

亚马逊云科技中国生态系统事业部总经理沈涛介绍说：“亚马逊云科技伙伴的解决方案一应俱全。全球有超过十万家的生态链合作伙伴网络。全球 90% 以上的财富世界 100 强公司，都在使用亚马逊云科技生态链合作伙伴提供的解决方案和服务。在中国，我们已经发展了数以千计的生态链合作伙伴，覆盖咨询合作、系统集成商、独立软件开发商、托管服务提供商，以及增值分销商等各种类型。”。从以上对比可以看出，虽然生态链合作伙伴网络在中国已有成效，但是面对百倍的数量差距中国区的生态链网络建设还需要更进一步。

4.4.2 SageMaker 应用市场目前没有在中国落地

SageMaker 的定位是帮助客户通过机器学习云平台更加快速的完成工程化，而非提供算法。专业的工作要交由专业的团队或公司，对于算法有定制化需求同时自身没有算法研究能力的客户，只能寻求 MarketPlace 的帮助，但该功能目前在中国区尚不可用。从下图 4.18 和 4.19 在应用市场的搜索结果可以看到，海外的 MarketPlace 可以提供 926 个算法跟模型来弥补 SageMaker 自身算法的不足，但在中国，并没有提供以 SageMaker 为交付方式的算法模型。

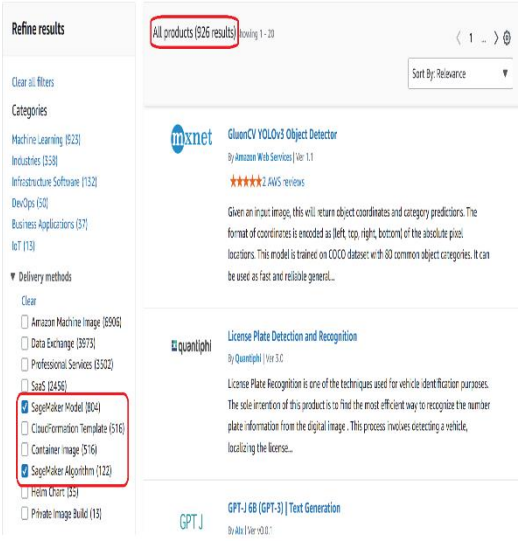


图 4.18 海外区 MarketPlace 的交付方式

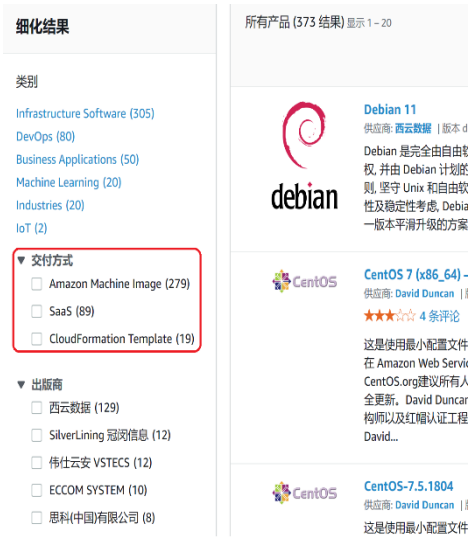


图 4.19 中国区 MarketPlace 的交付方式

4.5 运营总体评价

综上所述，SageMaker 在国内的运营能力存在几个比较突出的短板。整体上讲，技术运营方面缺少足够丰富多样的功能满足一般场景特别是中国本地化

场景的需求，价格运营上缺少性价比较高的新计算类型持续降低用户的总体拥有成本，品牌运营方面缺少中国明星企业的示范效应，生态链建设方面缺少应用市场的本地化接入。

5 亚马逊机器学习云平台 SageMaker 在中国运营策略优化

SageMaker 因其全球的技术领先优势，多措并举的价格运营策略，跨领域的品牌影响力，丰富的生态链网络，在进入中国后结合国内有利的宏观环境，采取了 OS 的增长型运营策略。但是随着国际逆全球化的趋势加剧，美元汇率的波动放大，SageMaker 在国内的运营策略需要适时地做出优化调整。结合上文运营问题的分析，SageMaker 需要将运营策略调整为 OW，即弱点强化的运营策略。通过抓住机会弥补短板、强化弱点，达到提升竞争力的目的。

5.1 技术运营策略优化

5.1.1 加速功能落地速度，全面补齐短板

从 SageMaker 的技术运营角度来说，加速其功能的补足，弥补其海外跟中国的不同体验，才能实现在中国运营三驾马车并驾齐驱的战略。在前述的分析中，SageMaker 2022 年新功能的平均延迟有 46.8 天，而更大的问题是如何解决 79 个新功能只有 10 个落地在中国的巨大鸿沟。从具体实施角度上分析，所有在中国运营的数据中心必须有严格的安全等级要求，其中工信部，发改委都有对应的准入标准，其目的也是进一步强化国家对数据的安全监管。随着监管的力度的增加，相信在未来对新功能引进会有更高的要求，但是这也是对全球处于机器学习云平台领导者 SageMaker 的一次进化自身运营能力的宝贵机会。对此，SageMaker 在中国的技术运营需要分析其技术功能的上下依赖，同时跟中国运营机器学习平台的相关法律法规结合，通过合理量化优先级利用精益生产的流水线方式，不断迭代更新，在加速短板提升的同时巩固优势功能的场景化落地。对于本土云厂商，如何让自己的机器学习云平台服务输出到海外，满足欧洲、北美对安全审计的要求，是值得其深入思考，并提前布局的。尤其是欧美等发达国家有相对完善的行业标准，在这点 SageMaker 在海外的运营经验值得国内云厂商借鉴。

5.1.2 丰富缺失服务的替代方案

对于 2022 年 69 个无法在国内落地的新功能，提供替代解决方案。这是解决方案架构师，机器学习专家，SageMaker 产品经理的主要职责。根据客户的

诉求，迅速整合资源，完成原型验证，以开源的方式分享给有需求的客户。同时配合技术引进，并在技术引进之后迅速触达客户，帮助客户针对业务场景做新功能的适配，完成一个闭环的技术管理。这对亚马逊云科技的人力资源，物理资源，需求管理提出了运营挑战。这也对技术运营中产品的全生命周期管理提出了更高要求。这种全周期的技术运营管理可以借鉴精益生产理论在丰田流水线的应用。流水线的应用从生产性行业扩展到机器学习平台开发领域需要做相关的适配工作。具体来说，首先，针对每一个缺失的服务做场景化调研，从对客户影响程度的角度出发，定义缺失服务的需求程度优先级。然后，成立小规模以研发为目的的技术团队，根据需求程度优先级来评估技术实现难度，其目的是从工程化角度推进关键需求功能的快速上线。最后，利用中国相对欧美较低的人力成本完成本土化工具、功能的开发，加速满足本土客户的通用需求。对于小众需求，该开发团队可以深入客户现场，有偿的在 SageMaker 的通用功能基础上做增值服务的二次迭代开发。

5.1.3 提高机器学习云服务本土化元素支持能力

对中国本土客户而言本土元素主要包括含中文的文字，视频，图片，语音等媒体资料。这些资料是机器学习场景的典型输入数据，也非常契合人工智能在真实应用场景的落地。提供以上功能的本地化解决方案可以有效弥补新功能引入不足的问题。如果可以满足更高级的需求如将媒资资源相互转换，会更加吸引中国客户在机器学习云平台的深度使用。以当下生成式 AI 的发展来看，中文文字翻译、中文语音翻译、中文文字跟语音的互转、视频提取字幕等高阶多模态的人工智能的应用门槛进一步降低，而实现其商用会有效帮助中国本土客户在其业务场景下的创新。

5.1.4 提供免费试用，培训，减小学习曲线

每个客户的场景，技术能力水平都不完全相同，必须提供在线培训，免费动手实验加速客户对产品的理解。这对缓解在技术运营分析中提到的陡峭学习曲线问题有所帮助。亚马逊云科技的技术博客所介绍的方案有极高的可复制性，其原因是每篇技术博客都有严格的文档审核流程，保证文档内容浅显易

懂，并且有详细的操作步骤。用户可以据此详细操作步骤体验解决方案的便利，同时省去花费大量时间阅读冗长的用户指南文档。但是目前这些技术博客通常以解决单个具体问题的形式存在，对于刚刚入门 SageMaker 的用户来说，仍然面临不知如何开始的问题，这就需要将这些散落的技术博客做统一整理，以一个端对端的形式组织起来，融入场景化需求，进一步加强对初级用户的入门级引导。对于成规模的初级用户，可以利用该端对端的学习资料组织免费培训。对于进阶客户，可以对新功能开展基于时间跟花费额度的免费试用策略。例如，SageMaker 推出了异步推理功能，进阶客户可以提出免费试用申请，亚马逊收到申请后，审核客户过往在 SageMaker 的使用情况以及潜在的场景匹配可能性给予申请者适当的补贴。而针对高阶客户，在功能开发阶段就引入他们的声音，邀请其作为白名单客户提前内测，提供反馈意见。功能正式上线后，对高阶客户做第一时间的技术博客跟进，以期通过成功的客户故事加强其他用户对 SageMaker 的使用信心。

5.1.5 加强销售人员培训力度，提高销售服务质量

虽然云平台可以降低机器学习的入门门槛，但是对于大部分零基础的客户来说，上手还是比较困难的。这就要求亚马逊云科技的一线员工了解 SageMaker 的基本功能，才可以发掘客户的深度需求，转化为商业机会。掌握机器学习服务的学习成本较高这一规律也同样作用于云厂商自己的员工。所谓一线员工是指所有对接客户需求的技术人员以及商务拓展人员，包括解决方案架构师，技术客户经理，商务经理，客户成功经理等。对于这些一线员工，需要掌握机器学习的一般方法论，SageMaker 的基本用法，才能在适当的场景帮助到客户，让丰富的替代方案，本地化的解决方案，新功能的引进变得有意义。与此同时，也可以对需求做到正向反馈，使需求能够被准确的传达，新方案可以被无缝的适配，做到基于客户真实场景的正向可持续性发展。

5.2 价格运营策略优化

5.2.1 构建健康，总体价格下降的价格体系

企业在做技术选型之前都会考虑整体性价比。以三年期 SageMaker 的整体

拥有成本为例，其计算跟存储的费用占到总成本的 92%以上。所以基础服务的价格体系决定了 SageMaker 的价格运营能力。Nvidia 的高端芯片禁止出口对于目前急需 GPU 算力的机器学习场景影响较大，如何保持合规合法的前提下满足国内对高端芯片的需求显得格外重要。首先，亚马逊云科技的破局之路是输出自研的性价比更高的推理和训练芯片。目前 Inf1 芯片已经在中国上线，可以弥补部分对高端芯片的缺口，但是目前还没有最新一代的训练芯片获批引入中国。其次，对于性价比较高的底层基础服务，优先引进。越是底层的服务其对上层的价格影响越大。最后，做好价格的管控。由于亚马逊云科技的定价体系参考其海外区，而且直接跟美元挂钩，面临 2022 年整体美元强势的背景，资源价格相对上涨。为了能给客户一个相对友好的价格，亚马逊云科技可以通过在固定周期内取人民币汇率的最低点作为汇率结算价格，减少汇率波动带来的价格变动，与此同时配合海外的降价策略，帮助中国客户维持一个总体下降的健康价格体系。

5.2.2 通过灵活补贴，更多免费试用，降低特殊客户的总体负担

SageMaker 的收费完全体现了按需付费的灵活性，但是每个客户的应用场景会有不同，如有的客户需要海量数据，但是模型简单，属于重存储轻计算的场景，而有的客户正好相反，重度依赖少量的数据，做大模型的训练。如果可以打通计算，存储单独计费的规则，提供一个长期的整体打包优惠措施供客户选择，那么对打算与亚马逊云科技有进行长期合作的客户来说会更友好。当客户承诺一个长期，高消费的用量，亚马逊云科技通过减少长期的风险成本给客户一个整体的价格优惠是一个双赢的合作。这种模式也是亚马逊云科技首创的。早在 2010 年亚马逊云科技就推出了预留实例功能，当客户承诺用指定的机型满一年或三年时，就会享受到最高 75%的优惠。而此处所建议的创新在于打通不同基础服务之间的通用折扣机制。在 SageMaker 的使用场景下，建议对 EC2，S3，EFS 实施跨服务的分阶梯打包促销策略，帮助客户进一步基于长期合作优化成本。

5.3 品牌运营策略优化

关于品牌的影响很难有一个定量的分析，而且不同部门所感知到的影响因

素各不相同，为了更好的发散想法，找到真正可能会影响品牌运营的主要原因提高集体决策的效率，研究者召开了头脑风暴。在头脑风暴的会议中一共有七位来自不同部门的同事参加，总共收集到三十种不同的看法。经过头脑风暴环节的开放性问题提出，需要对问题进行总结、概括，才能达到以有限的资源投入改进主要问题的目标。在基于鱼骨图的分析方法中，作者根据问题影响的大小和根因的归纳，提炼为如下五个方面的六个核心问题,如下图 5.1 所示。基于核心问题的提出，作者通过以下四个方面的措施加以覆盖解决。头脑风暴的开展与鱼骨图的绘制详情请参见附录 2。

图 5.1 导致品牌形象不好的鱼骨图表示

5.3.1 加强中国头部标杆客户宣传力度

随着中国机器学习产业近些年的迅速崛起，中国也有一批明星企业屹立于不同行业的科技公司之巅，如社交类的抖音，短视频的快手。还有迅速崛起的其他新兴行业的独角兽公司，如中国的 AI 四小龙，中国新能源汽车的蔚、小、理（蔚来，小鹏，理想汽车）。这些公司在机器学习平台的带动能力比海外的传统公司要强。另外所有机器学习项目有一个共同特点是失败概率高，试错成本高，这导致一旦有了零到一的成功，就可以相对容易的做到一到一百的复

制，而这正是中国企业所擅长的。所以对于有亮点的明星企业在 SageMaker 的成功案例分享，是可以带动 SageMaker 在中国的更大范围的使用。在品牌运营的分析中，研究者分析了技术博客，旗舰峰会和全球 Re: Invent 大会。以上三个已有平台都可以打造品牌知名度，充分利用好这些媒资平台可以有效帮助 SageMaker 扩大其行业影响力。

5.3.2 突出亚马逊云的品牌价值营销

目前通过官网能够找到的正在使用 SageMaker 的中国客户只有大宇无限和 OPPO，显示出 SageMaker 对中国客户的渗透率较低。如何对中国客户提高 SageMaker 的曝光率以及触及率是品牌运营应该考虑的首要问题。作为世界电商的领袖亚马逊具有较高的品牌价值，而亚马逊云科技就是亚马逊电商在运营 IT 系统时遇到棘手问题后所提出的外租资源的解决方案。电商的机器学习场景大多数跟个性化推荐，客户精准画像分析有关。在 SageMaker 的品牌运营中，可以通过宣传 SageMaker 如何帮助亚马逊电商解决其客户的个性化需求的案例，起到破冰电商客户的效果，从客户场景出发，直击客户痛点，提高在电商市场的认知率。再利用电商客户的头部明星企业做其他行业的从零到一的市场培育，有机会全面影响其他行业客户。

5.3.3 做行业化市场活动，做最懂客户行业的解决方案

目前的云计算市场已不再是当初提供几个独立的云服务，满足服务的高可用，按需付费的简单场景了。而是深入到各个领域，行业，做行业的整体解决方案。这种行业属性一般都伴随者区域性特点，例如，对于汽车行业，各个国家的行业标准或许不同，对于数据的敏感度划分，审计要求也不一样。这就要求云厂商不能只关注自身服务，而要关注上层应用的行业规范。这就要求云平台提供商必须从之前的纯技术管理宣传理念，转换成面向客户行业的基于客户品牌的运营方案。还以汽车行业举例，车企属于传统制造业，面对 IT 技术的变革并不知道该如何多、快、好、省的上云。作为云厂商首先要了解客户上云的诉求，做场景化的功能划分，如了解客户是属于车联网的低延时高安全性需求，还是自动驾驶的高机器学习算法模型的开发场景，亦或是智能座舱的高交

互式 AI 服务体验要求。这对云厂商来说需要在自身的组织架构上做出调整，将更多的品牌运营策略融合于特定行业。因此，基于 SageMaker 的行业性方案，市场活动是提高其市场占有率以及品牌效应的有利手段。

5.3.4 SageMaker 走进高校，赞助科研机构

机器学习的理想试验场和加速器是高校和科研院所，因为其代表着前沿技术研究方向和不断的人才输出。有人才作为基础结合成功的科研结果，SageMaker 在工程项目中的影响力也会日益增大。另外有了科研院所的背书也更容易让国有企业、事业单位认可。目前国际上比较多的方式是通过与研究机构共同发表学术论文对自身的品牌做对外宣传。如 Google 跟多伦多大学在大语言模型中对基于 Transformer 的多注意力模型的研究，开创了大语言模型的新纪元。也间接地对 OpenAI，Anthropic 这样的独角兽公司的诞生产生了积极的影响。

5.4 生态链运营策略优化

5.4.1 持续扩大开源技术社区影响力

开源社区对学术的影响力也越来越大，因为很多新算法的提出都是基于前人的研究，并融入了创新性的思考。2018 年亚马逊云科技在上海成立了 AI 研究院，可以通过研究院的学术能力跟高校合作，共同推进算法的研究同时贡献开源社区，在学术界持续施加影响。Github 是开源社区最具影响力的门户，随着系统的微服务化架构走向普及，代表着小而美设计哲学的应用系统大行其道。更多的系统组件也变得越来越插件化，使得新功能的适配越来越容易。比如对系统日志的监控，对流量进出的规则管理，对细粒度安全权限的管控等。这样的微服务架构也对丰富 SageMaker 生态起到巨大的推动作用。做好 SageMaker 在开源社区的插件开发，更方便的适配新功能，争取在生态上取得更大的影响力。

5.4.2 加速 SageMaker 应用商店在中国落地

海外针对 SageMaker 的模型和算法在其应用市场有 900 多个可选产品，涵

盖了各个行业的各种数据类型。而中国区目前还不支持模型跟算法的第三方应用商店。如果该功能落地中国会在很大程度上缓解 SageMaker 中国区的功能性不足。除此之外，使用国内 SageMaker 的客户也可以把自己的算法，模型反向输出给中国 SageMaker 的其他客户，促进细分领域跟平台普适性模型的算法互补。MarketPlace 是另一个技术社区，但更偏向于技术变现。利用 SageMaker 机器学习云平台，客户可以把在 SageMaker 优化后的模型或算法有偿出让给其他 SageMaker 客户，达到商业盈利的目的。

5.4.3 推进大数据，机器学习领域生态链合作伙伴建设反哺海外区

中国在世界人工智能领域有较大影响力，从学术论文的发表量到被引用次数都位列世界前列。尤其是自动驾驶领域，在这个集汽车工程，高端制造，系统集成，机器学习算法融合的复杂行业孕育了如 Momenta，轻舟，文远这样的国际优秀企业。而这些公司都是亚马逊云科技中国区的重要企业级客户。他们在模型训练，算法能力上处于世界领先地位。完全可以通过 SageMaker 平台对海外客户尤其是海外主机厂做基于自动驾驶场景的解决方案价值输出，从而反哺海外的生态，做到国内、国外的全生态链融合发展。

6 结论

6.1 研究结论

本研究的目的是如何提高 SageMaker 在中国运营能力。将 SageMaker 在中国和海外的运营策略作为研究对象。分别从 SageMaker 的技术体系框架、价格构成，品牌运营模式和生态链运营策略进行多维度的综合分析。本文提出的机器学习云平台运营能力分析框架可以被其他云厂商借鉴，用来评估自己的机器学习云平台运营能力，发现自身运营瓶颈。SageMaker 在中国运营的优化措施主要包括以下几个主要方面。

首先在技术上，加速中国区功能落地速度，全面补齐短板可以显著减少由于缺失功能而导致的服务可用性挑战。通过丰富缺失服务的替代方案来缓解目前 SageMaker 在国内的技术能力不足。提高机器学习云服务在本土化元素的支持，满足中国客户的个性化需求。提供免费试用，培训，减小学习曲线的同时加强自身一线员工的培训力度，提高其技术服务质量。

在价格运营上，要构建稳定、且持续走低的价格体系，面向不同类型场景采取不同价格策略，如通过直接的降价、上线高性价比的新一代替代产品，阶梯消费等价格策略吸引新客户。通过灵活补贴，更多免费试用，降低特殊客户的总体负担。

在品牌运营方面，加强中国头部客户标杆示范性的宣传力度。突出亚马逊云计算的品牌价值，影响更多电商行业客户。做行业化的品牌运营，做最懂客户行业的解决方案，帮助客户从业务上解决场景化问题。推进 SageMaker 走进高校，赞助科研机构持续扩大影响力。

在生态链建设上，持续扩大开源技术社区影响力，加速 SageMaker 应用商店的加速落地。推进大数据，机器学习领域生态链合作伙伴建设反哺海外，做健康的闭环生态。

6.2 研究局限性

由于学术界对于机器学习云平台运营策略的研究较少，可查询的文献资料

比较有限，因此本研究尚存一些不足之处。

首先，由于亚马逊公司在中国经营的保密性要求，研究者无法获得关于该公司产品的市场销售数据以及财务数据，仅能根据其网站，第三方的调查报告所公开的资料进行梳理、分析，在部分内容上可能还存在介绍不够详细的情况。

其次，由于本研究立足于管理视角的剖析，关于机器学习技术性的分析多以通俗文字带过，对于机器学习技术本身的学术性探索不够深入。

最后，随着生成式 AI 的持续火爆，机器学习云平台的发展与演进也会加速推进。本研究大部分统计数据截至于 2022 年 12 月底，所展开的论述跟结论可能有时间的局限性。

参考文献

[1] Barry Murphy.AWS Named a Leader in 2022 Gartner Magic Quadrant for Cloud AI Developer Services [Z]. July.2022.https://aws.amazon.com/cn/blogs/apn/aws-named-a-leader-in-2022-gartner-magic-quadrant-for-cloud-ai-developer-services.

[2] Frank Wan. 迎接未来, 砥砺前行——2022上半年中国人工智能市场份额发布 [Z]. International Data Corporation,Jan.13.2023.https://www.idc.com/getdoc.jsp?containerId=prCHC49990323.

[3] Rajkumar Buyya, Chee S-Y,Srikumar Venugopal et al.Cloud computing and emerging IT platforms:Vision, hype and reality for delivering computing as the 5th utility[J].Future Generation Computer Systems, Jun 2009,Volume 25,Issue 6,599-616.

[4] Michael Cusumano. Cloud computing and SaaS as new computing platforms[J]. Communications of the ACM, April 2010, Volume 53, Issue 4, 27–29.

[5] Cusumano, M. Technology strategy and management: The evolution of platform thinking[J]. Communications of the ACM, 2010, 53(1), 32-34.

[6] Wu, W. W, Lan, L. W, & Lee, Y. T. Exploring decisive factors affecting an organization's SaaS adoption[J]. International Journal of Information Management, 2011, 31(6), 556-563.

[7] Gupta, P, Seetharaman, A, & Raj, J. R. The usage and adoption of cloud computing by small and medium businesses[J]. International Journal of Information Management, 2013, 33(5), 861-874.

[8] Lehmann S, Buxmann P. Pricing strategies of software vendors[J]. Business & Information Systems Engineering, 2009, 1(6), 452-462.

[9] Márquez-Vera et al. Serverless machine learning: architectures[J]. platforms and applications, Applied Sciences, 2021, 11(7), 3097.

[10] Song, W, & Guo, J. Optimization principles for selecting machine learning platforms[J]. IEEE Access, 2020, 8, 191855-191868.

[11] Liao, S, Widjaja, A. E, & Wang, C. Brand equity and customer stickiness in AI cloud services[J]. Industrial Marketing Management, 2022, 103, 36-47.

[12]倪光南，陈江，李琦. 面向云计算的服务科学体系[J]. 计算机研究与发展 2012, 9, 1815-1831.

[13]Chen, Y., Lee, C., & Chou, T, Key success factors when Western cloud service providers enter the Chinese market: Case studies of Microsoft, Amazon, and Salesforce[J]. Journal of Global Information Management, 2020, 28(1), 282-308.

[14]叶劲松，吴翠萍. 关系规范、技术融合与平台生态系统协同演化[J]. 软件学报, 2022,33(04):1228-1246.

[15]陈念, 李国良, 王珺. 云服务定制化关键技术与应用[J]. 计算机研究与发展, 2018, 55(01):146-159.

[16]张晓，郑爽，周强. 基于消费者－品牌关系的公有云选择研究[J]. 情报理论与实践, 2017,40(07):99-104.

[17]吴江，章源. 基于 PEST 分析法的我国农业大数据发展环境研究[J]. 湖南农业大学学报(社会科学版)，2021,22(01):104-110.

[18]潘福利, 吴小川. 中国食品电子商务发展 PEST 分析[J]. 商业研究，2017(15):201-205.

[19]Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A. & Zaharia, M. A view of cloud computing[J]. Communications of the ACM, 2010, 53(4), 50-58.

[20]许志端, 李瑞鹏, 黄涛. 云计算业务模型及定价策略研究综述[J]. 计算机应用研究, 2013, 30(7):1828-1832.

[21]Hayes, B, What is Cloud computing[J]. Communications of the ACM, 2008, June, 51(7), 9-11.

[22]Sculley, D., Holt, G. Hidden technical debt in machine learning systems[J]. Advances in neural information processing systems, 2015, pp. 2503-2511.

[23]Fernández, A., Insfran, E., & Abrahão, S, Usability evaluation methods for the web: A systematic mapping study[J]. Information and Software Technology, 2018, 93, 149-162.

[24]F.Robert Jacobs,Ricard B.Chase. Operations And Supply Chain management 15th Edition[M]. 北京, 机械工业出版社, 2023.4 V 5-6, 23.

[25]王蒙征. 构建面向 21 世纪的中国经济发展理论[J]. 经济研究参考, 1999, 第 1 卷，第 10 期.

[26]赵蓓, 付会霞. 比较优势视角下我国电力工业发展战略研究[J]. 科技管理研究, 2010, 第 1 卷，第 12 期.

[27]王飞跃, 程若辰, 王科俊等. 数字孪生驱动的精益产业新模式创新[J]. 自动化学报, 2022, 第 48 卷, 第 4 期.

[28]张爱卿, 王秋实, 张丹. 基于精益生产理念的供应链管理模式创新研究[J]. 产业与科技论坛, 2020, 第 19 卷, 第 11 期.

[29]美通社. 2021-2022 年中国公有云市场现状及趋势研究报告[R].CCW Research,2022,9.26. https://www.prnasia.com/story/376883-1.shtml.

[30]Van Baker, Svetlana Sicular, Erick Brethenoux, Arun Batchu, Mike Fang, Magic Quadrant for Cloud AI Developer Services[R], 23 May 2022. https://www.gartner.com/en/documents/3997377.

[31]Paul Krill. The total cost of ownership of Amazon SageMaker[Z]. InfoWorld.com, 2021. https://pages.awscloud.com/NAMER-ln-GC-400-machine-learning-sagemaker-tco-learn.

[32]高增亮. Amazon SageMaker 产品能力评级报告-2021 年第三季度[Z]. 字母点评，2022.12.12. https://p.shaqiu.cn/#/reportdetail?id=32.

后记

光阴似箭，日月如梭，三年的 MBA 研究生求学生活即将告一段落。回想这过往三年多的学习生活，面对即将告别的校园环境，心中无限感慨。回想起 2019 年研究生联考的备考，联考成绩发布后突如其来的疫情，以及这三年求学过程中疫情的反反复复，恰如人生的一个小小的缩影充满了未知与希望，好奇与挑战。我们经历了线上面试、授课、考试，感受了极其有时代特色的研究生校园生活。我们在 2020 年疫情初步清零时，有过一年珍贵的线下学习经历，让我的学校生活在工作十几年后得以延续，在此我特别感谢学校的领导，排除万难，在非正常的社会秩序下保障我们的学习生活，能够一步一个脚印的走到毕业答辩现场，感谢我可爱的同学们，他们来自全国各个省市，在分组讨论、案例分析时贡献着独特的视角。我还要感谢我的家人，在学业期间，持续给我鼓励与支持。最后感谢我的论文指导老师，帮助我在书写论文的同时系统地回顾了这三年所学的所有内容，并融入到本研究的写作中。

作者：朱煜

2023 年 8 月 24 日

附录 1：三位高管访谈提纲

采访主题 1：亚马逊云科技在中国运营战略

被采访人：亚马逊全球副总裁、亚马逊云科技大中华区执行董事张文翊

采访时间：2022-12-06

采访提纲：

1.请您介绍下目前国内云计算产业的发展现状？

2.您认为亚马逊云科技在国内发展的优势是什么？

3.您认为国内云厂商的主要优势是什么？

4.国家提出打造数字经济新优势，您认为中国数字经济进展如何？

5.亚马逊云科技如何助力企业提高数字经济质量？

6.您说的数字经济技术底座具体包含哪些方面？

7.作为技术底座之一的“拎包入住”的创新环境，指的是什么？

8. 技术底座中“按需付费”的高灵活性和成本优势是怎样做到的？

9. 您能展开描述下技术底座中“降低新技术使用门槛”的案例吗？

访谈主题 2：亚马逊云科技在生成式 AI 的布局

被采访人：亚马逊云科技大中华区机器学习产品总监张洋

采访时间：2023-04-16

采访提纲：

1.您认为今年崛起的生成式 AI 对企业来说会有什么样的影响？

2.生成式 AI 的兴起对亚马逊云科技的运营来说，有哪几方面的影响？

3.您可以具体解释下您说的技术运营所涉及的问题吗？

4.您为什么认为成本运营是非常重要的运营因素？

5. 亚马逊以及亚马逊云已经很有知名度，为什么品牌运营仍非常重要？

6. 为什么生态链建设尤其是本土化生态链建设也同等重要？

采访主题 3：亚马逊云科技如何深耕中国市场

被采访人：亚马逊云科技中国国际客户及生态合作伙伴，生态系统事业部总经理沈涛

采访时间：2021-11-16

采访提纲：

1. 您在亚马逊云科技主要负责中国的跨国企业与生态链合作伙伴业务。这个业务部主要负责的内容是什么？

2. 亚马逊云科技为什么重视跨国企业业务？

3. 跨国企业在中国的分公司要落地一个全球项目时，面对中国不一样的环境，需要有自己的创新，亚马逊云科技是如何帮助它们的？

4. 跨国企业在中国面临着机遇的同时，也面临着数字化转型的压力。与本土客户相比，跨国企业在中国实现数字化转型有哪些独特需求和挑战？

5. 本土团队对本土市场需求有着独特见解，全球团队则有更多技术和经验，亚马逊云科技的这两个团队是如何合作，共同服务于跨国客户的？

6. 在服务跨国企业的过程中，生态链合作伙伴扮演着怎样的角色？

7. 亚马逊云科技在哪些行业积累了丰富成熟的经验？

8. 除了汽车与医疗行业，亚马逊云科技还有哪些行业跨国企业的客户？

9. 整个汽车行业面临着“新四化”机遇。亚马逊云科技是如何赋能汽车行业客户，抓住这些机遇的？

10. 在中国，亚马逊云科技赋能跨国企业未来发展的重点和方向是什么？

附录 2：头脑风暴 - 对品牌运营关键影响因素分析

1、会前准备阶段

（1）确定头脑风暴的议题和任务目标为：收集对 SageMaker 品牌运营的关键影响因素

（2）会议组织形式：邀请来自市场，研发，人事，运维，法务，财务代表，共计 7 人，进行头脑风暴法。作者为主持人，只负责主持会议，不评论任何的设想。1 个记录员，可适当配以录音、摄像等方式来确保完整记录每一个设想。要求与会者善于想象，语言表达能力强。

（3）会议准备：配备白板，投影仪，以及记录用的纸、笔和笔记本电脑等。记录员可根据情况选择使用传统的纸和笔做记录，也可以直接在电脑上用思维导图软件进行记录。

（4）会议环境：可以容纳 10 人的会议室。

（5）会议时间：2023.5.5

2、开放讨论阶段：

(1) 主持人宣读头脑风暴法的原则：

1、自由发散原则：营造轻松氛围，让所有与会者放松思想，不受任何约束限制，能从不同角度、层次、方向来自由思考，体长大胆设想，力求标新立异的创新创意出现。

2、主题聚焦原则：头脑风暴的目的是要追求尽可能多的想法，思维可以发散，天马行空般地随意设想，但必须始终围绕所讨论的主题，不能偏离主题。

3、以量求质原则：头脑风暴鼓励多思考，多做设想，设想越多，创造性、创新性的设想就可能越多。

4、延时评判原则：在进行头脑风暴的过程中，应当集中精力开发创造性的想法，不对别人的任何想法做任何的评价，既不肯定也不否定，也不必过分自谦。

5、二次创新原则：相关部门的主管或决策者应当认真阅读头脑风暴的成果文档，必要时进行更高层次的头脑风暴讨论。

(2) 主持人组织会后信息整理得出以下因素

- 品牌定位不明确,使消费者对品牌印象混乱
- 产品和服务质量参差不齐,辜负消费者期望
- 说明文档以及免费学习资料较少
- 虚假夸大产品,被曝光后严重损害品牌信誉
- 对 AI 等比较前沿科技熟悉的商务人数较少
- 盲目广告轰炸,给消费者留下负面印象
- 价格策略频繁改变,给消费者造成焦虑或不信任
- 售后服务差,客户投诉处理不当,导致客户流失
- 企业丑闻不断,负面新闻伤害品牌形象
- 极少头部标杆客户的站台
- 产品升级过快,使老客户感到被抛弃
- 设计风格频繁大幅更改,难以建立持久印象
- 品牌文化模糊,员工行为表现不统一
- 缺失对未来人才的培养
- 社会责任意识淡薄,环境和道德风险存在
- 品牌拓展过度,新品种严重扩散品牌认知
- 竞争对手品牌崛起,本品牌优势被弱化
- 管理层频繁更换,品牌发展战略难以持续
- 没有形成合力,在一个领域打开局面，作为宣传热点
- 数字化转型不力,难以跟上新消费趋势
- 消费群体变迁,品牌定位和传播失去吸引力
- 中国客户在国际区的声音太小
- 社交媒体营销差强人意,很难产生用户兴趣和参与
- 新品牌太相似,难以区分,分散了注意力
- 品牌创新不足,一成不变,渐趋老牌化
- 未按消费心理调整营销策略,沿用传统方式

3、归纳主要问题阶段 – 鱼骨图绘制：

经过头脑风暴环节的开放性问题提出，需要对问题进行总结、概括，才能达到以有限的资源改进主要问题的目标。根据问题影响的大小和根因的归纳，提炼为如下五个方面的六个核心问题。

- 说明文档以及免费学习资料较少
- 对 AI 等比较前沿科技熟悉的商务人数较少
- 极少头部标杆客户的站台
- 缺失对未来人才的培养
- 没有形成合力,在一个领域打开局面，作为宣传热点
- 中国客户在国际区的声音太小

综合五个方面，6 个核心问题和以上所有的问题列表，鱼骨图绘制如下：