通用软件初创企业与机器学习初创企业的软件工程实践分析

Bishal Lakha计算机科学系
博伊西州立大学博伊西，爱达荷州，美国
bis
hallakha@u.boisestate.edu

Kalyan Bhetwal计算机科学系
博伊西州立大学博伊西，爱达荷州，美国
kalyanbhet
wal@u.boisestate.edu

Nasir U. Eisty计算机科学系 博伊西州立大学博伊西，爱达荷州，美国
nasireisty@boisestate.edu

arXiv:2304.01523v1 [cs.SE] 4 Apr 2023

摘要——背景：**除了初创软件公司在采用适当的软件工程实践时面临的固有挑战外，机器学习技术具有非确定性，这使机器学习初创企业（ML 初创企业）面临更大的困难。**目标：**因此，本研究旨在全面了解机器学习初创企业所采用的软件工程实践，并识别其额外需求。**方法：**为实现这一目标，我们对过去 21 年发表的 37 篇论文进行了系统性文献综述，所选论文涵盖通用软件初创企业和机器学习初创企业。我们收集数据，以了解软件开发生命周期五个阶段中的软件工程实践：需求工程、设计、开发、质量保证和部署。**结果：**我们发现，机器学习初创企业与通用软件初创企业在软件工程实践方面存在一些显著差异，其中最突出的是数据管理和模型学习阶段。**结论：**尽管机器学习初创企业面临许多与通用软件初创企业相似的挑战，但使用随机机器学习模型带来的额外困难，要求它们采用不同的软件工程实践策略，以开发高质量产品。

索引词—软件工程、机器学习初创企业、软件初创企业、系统性文献综述

一、引言

机器学习正日益广泛地应用于各种软件应用中。初创企业和小型公司正积极采用这项技术。它们是为不同领域实施创新且最先进的机器学习解决方案的引领者 [49]。机器学习应用的挑战性，以及初创企业的局限和特殊性，导致其采用了略有不同的软件工程实践。然而，这些系统的开发、部署和维护仍然缺乏最佳实践。

由于机器学习具有非确定性，机器学习系统的各个软件工程方面都变得复杂 [24]。此外，缺乏用于测试机器学习系统的适当且成熟的工具 [24]。因此，将机器学习集成到软件应用中，促使各组织不断改进其开发流程 [3]。Amershi 等人 [3], 开展的一项研究考察了微软不同软件工程团队开发人工智能（AI）产品时所采用的方法。研究表明，人工智能组件本质上比其他软件组件更复杂；集成这些组件时，可能会出现非单调错误行为。他们提出了端到端流水线支持、数据收集、清洗和管理工具，重点关注编程缺陷和模型缺陷等问题，以缓解此类问题。近期，少数公司采用了机器学习运维等新方法，但仍面临许多挑战 [52]。这些挑战在初创企业中尤为突出。

超过 60% 的初创企业未能存续五年，风险投资家支持的初创企业中有 75% 以失败告终 [40]，而超过 90% 的初创企业最终破产[20]。造成这种情况的一个重要原因是软件工程实践不当，导致产品存在缺陷[14]。此外，许多初创企业开发最小可行产品需要一年以上，而产品交付延迟会带来额外的财务负担，最终导致初创企业失败 [32]。同样，组织未能积极吸纳开发人员参与，也会导致失败 [1]。这些现象同样是糟糕的软件工程实践造成的。

软件工程中的不良实践并不总会导致初创企业失败。其他因素还包括不同类型的技术债务，例如代码债务、架构债务或测试债务。这些债务可能迅速累积，影响初创企业的绩效和发展 [33]。人工智能/机器学习初创企业比传统初创企业更容易失败，上述原因也会导致它们失败。因此，了解它们采用了哪些软件工程实践，以及哪些最佳实践至关重要。

关于科技行业软件工程实践的研究已有不少。此外，也有许多研究探讨初创企业中的软件工程实践。然而，针对机器学习初创企业及其软件工程实践的研究寥寥无几。这一研究空白促使我们将研究重点放在该领域。此外，我们意识到，机器学习领域正在迅速发展，因此有必要研究这些公司，尤其是初创企业中的软件工程实践变化，并将其与通用软件初创企业进行比较。

因此，我们的研究主要目标如下：

- 识别软件初创企业普遍遵循的关键软件工程实践
- 识别机器学习初创企业所遵循的具体软件工程实践

II. 背景

智能手机、平板电脑、笔记本电脑、智能手表等各类电子设备的出现，推动了软件行业的蓬勃发展，催生了大量软件初创企业 [9]。深度学习作为机器学习的一个子领域，在广泛应用中取得成功，促使多家公司采用机器学习，并推动初创企业迅速发展 [44]，为经济贡献了数十亿美元 [12]。

通用软件初创企业。软件初创企业是区别于传统成熟企业的中小型企业，它们专注于在有限的时间和资源内开发创新产品 [55]。由于自身特点，它们面临诸多挑战，例如组织管理经验不足、缺乏财务和人力资源，以及受到投资者、客户、合作伙伴等多方影响，还需要持续开发动态颠覆性技术 [50]。

机器学习初创企业。机器学习初创企业提供语音识别、视频分析等不同的机器学习服务，或将机器学习作为其产品的一部分。医疗健康 [22] [56], 金融 [57],时尚 [35], 等不同领域的许多初创企业也在其产品中采用人工智能和机器学习。在开发此类产品时，初创企业会因研究与开发之间的差距而面临诸多不确定性 [45]。由于软件初创企业更容易失败 [10], , 这些不确定性使机器学习初创企业更为脆弱。然而，良好的软件工程实践所带来的协作、合作与开放，有助于此类初创企业取得成功 [8]。

A. 研究问题

为实现我们的目标，我们提出了两个研究问题，并对不同渠道发表的多篇论文进行了系统性文献综述。

研究问题1：通用软件初创企业采用了哪些软件工程实践?

通过研究问题1，我们旨在了解通用软件初创企业采用了哪些软件工程实践，以及这些实践与传统软件工程实践有何异同。这些实践将展现软件初创企业的特点，并说明它们选择特定软件设计模式和开发实践的原因。

研究问题2：机器学习初创企业还采用了哪些额外的软件工程实践?

通过研究问题2，我们旨在根据软件工程实践区分通用初创企业与机器学习初创企业。我们还计划了解机器学习产品开发涉及的不同步骤、哪些通用实践仍然有效，以及为应对这些步骤需要进行哪些调整、改进和创新。

III. 研究方法

我们收集了已发表的论文，并逐篇人工审阅所有选定的论文以收集数据。为保证研究质量，第一作者和第二作者分别独立审阅了每篇论文。收集数据后，我们对数据进行了分析和综合，以回答研究问题。本节将详细介绍我们的研究方法。

A. 纳入与排除标准

我们仅纳入以英语发表的论文，并将通用软件初创企业相关的期刊论文、会议论文或研讨会论文集纳入考虑。由于关于机器学习初创企业的已发表论文较少，我们纳入了少量博客。我们排除了视频、播客、新闻文章、访谈等其他灰色文献。此外，在本系统性文献综述中，我们仅考虑截至2021年11月发表的论文。

B. 数据库检索

我们拟定了检索字符串，以下载符合研究目标的论文。随后，我们使用拟定的检索字符串在 IEEE Xplore、ACM Digital Library 和 Google Scholar 中检索论文。起初，我们在这些在线数据库中使用了两个检索字符串。查询 1 为“软件工程与初创企业”。查询 2 为“机器学习与软件工程及初创企业”。检索结果数量很多，但只有少数相关。我们进一步在 Google Scholar 上检索，以丰富我们的论文数据库。

我们进一步根据通用软件初创企业的软件开发生命周期（SDLC）各阶段制定了子查询。在查询 3 中，我们搜索了“软件工程与初创企业与需求分析”。在查询 4 中，我们搜索了“软件工程与初创企业与设计”。在查询 5 中，我们搜索了“软件工程与初创企业与实现与 sdlc”。在查询 6 中，我们搜索了“软件工程与初创企业与测试与 sdlc”。在查询 7 中，我们搜索了“初创企业与部署与 sdlc”。随后，我们下载了元数据以供进一步处理。

我们的目标是收集基于软件开发生命周期的机器学习初创企业相关论文；我们在 IEEE Xplore 和 ACM Digital Library 中进行了类似的查询搜索，但没有找到相关论文。随后，我们使用“深度学习”和“人工智能”等机器学习的其他术语来丰富机器学习初创企业数据库，并找到了一些额外的有用论文。

C. 元数据收集

我们通过数据库检索，将论文的元数据（标题、作者姓名、摘要、发表年份、URL、引文等）收集到一个 CSV 文件中，以便识别重复论文并进行验证。这些数据也便于我们根据摘要手动筛选论文。IEEE Explorer 内置了可将所有搜索结果的元数据导出为 CSV 文件的功能，但 ACM Digital Library 和 Google Scholar 不具备这些功能。因此，我们使用 Python 库 BeautifulSoup 构建了一个网页抓取工具，并从 ACM Digital Library 的搜索结果中提取了元数据。

D. 去重与验证

我们从 IEEE Explorer 收集了 389 篇关于通用软件初创企业的论文元数据，从 ACM Digital Library 收集了 12,274 篇，从 Google Scholar 收集了 9,640 篇。同样，我们从 IEEE Explorer 收集了 13 篇关于机器学习初创企业的论文，从 ACM Digital Library 收集了 10,040 篇，从 Google Scholar 收集了 2,380 篇。我们还收集了网站 ml-ops.org 上整理的 39 篇论文。随后，我们根据元数据从列表中删除了所有重复论文。我们使用名为 Pandas 的 Python 软件包完成了这项工作。

我们还注意到，从检索中收集的许多论文的摘要并未包含用于检索的关键词。为了剔除这类论文，我们使用了正则表达式。正则表达式使用一系列字符来查找文本中的特定模式。我们利用它检查摘要中是否包含所有必需的关键词。如果在摘要中未找到一个或多个关键词，我们就会将该论文从列表中删除。图 1 展示了在根据关键词是否存在进行筛选前后，属于通用软件初创企业不同软件开发生命周期的论文数量。

E. 雪球抽样

我们针对通用软件初创企业进行的数据库检索获得了足够数量的论文，能够满足我们的研究需求。然而，我们针对机器学习初创企业获得的论文非常少。因此，我们使用雪球抽样技术来扩大机器学习初创企业的数据库。我们查阅了所找到论文中的引文，并根据纳入与排除标准检查其中是否有与我们研究相关的论文。此前我们有 16 篇相关论文，经过雪球抽样后，获得了 21 篇论文。

F. 人工筛选与最终确定

去重和验证后，我们共有 92 篇论文，其中 72 篇来自通用软件初创企业，20 篇来自机器学习初创企业。自动筛选有效地剔除了无关论文，但有几篇论文虽然包含所需关键词，却对我们的研究没有帮助。因此，我们阅读了所有剩余论文的摘要，并进行了人工筛选。最终，我们选出 37 篇论文，其中 27 篇来自通用软件初创企业，10 篇来自机器学习初创企业，并将它们分配到软件开发生命周期的不同阶段，如表 I 所示。

表 I 通用软件和机器学习初创企业按软件开发生命周期阶段划分的最终论文数量

阶段	一般初创企业	机器学习初创企业
需求工程	5	3
设计	7	×
数据管理	×	4
开发	6	×
模型学习	×	4
测试	7	5
部署	2	5

IV. 结果

本节将按研究问题呈现我们的研究发现。

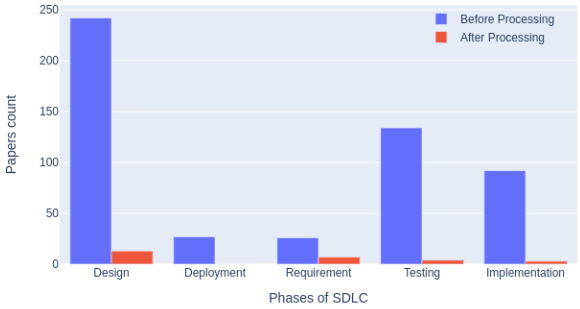


图 1. 筛选前后各软件开发生命周期阶段的论文数量

A. 研究问题1：通用软件初创企业遵循哪些软件工程实践？

我们根据软件开发生命周期，在表 II 中总结了通用软件初创企业遵循的各种实践，并在以下小节中进行了详细讨论。我们还讨论了研究中发现的潜在挑战和解决方案。

1) 需求工程：这是软件开发过程的第一部分。恰当定义的需求有助于更好地解决问题，并开发出质量最佳、效率高的软件。在软件开发生命周期的这一阶段，模糊的高层级用户需求会被转化为完整、精确且正式的规范，以供后续软件开发使用 [11]。这一过程涉及软件生产者与最终用户之间的互动。需求工程有多种方法。初创企业也会根据自身需求采用不同的技术。

大多数时候，初创企业会因软件需求表述不清而难以确定应该开发什么 [38]。Melegati et al. [38]采用扎根理论方法，就初创企业的需求工程实践，对来自不同巴西初创企业的创始人或经理共进行了九次访谈。他们发现，创始人或经理在选择初创企业的实践方式时发挥着重要作用。他们还观察到，软件开发中的需求工程成熟度各不相同。有些初创企业没有任何正式流程；另一些则有明确的需求工程步骤。在需求实践较为成熟的情况下，这通常主要是由市场压力或需求所致。需求工程因初创企业的性质而异，例如面向客户的初创企业或用户目标导向型初创企业。对于面向客户的初创企业，所有需求都由客户设定；对于用户目标导向型初创企业，需求则由所有者设定。两种情况下都会进行验证，以检查需求是否可行。这一阶段可能不会进行软件开发，或只进行少量开发。开发完成后，还会进行最终验证，以确认需求是否得到正确实现。

Rafiq et al. [43] 研究了全球范围内三家初创企业的需求获取过程。他们发现，初创企业的需求工程非常初级且非正式，并随着产品开发而不断发展。在需求工程流程开始之前，所有者心中已经有了初步构想。随着时间推移，这些构想逐渐成熟。最常见的需求获取技术包括开展访谈、原型设计和头脑风暴。此外，也会采用一些非常规方法，例如竞争对手分析、团队协作讨论和用户模型。

Alves et al. [2] 对初创企业中的需求工程进行了文献综述。他们发现，初创企业采用灵活且非正式的需求工程实践。此外，他们观察到，初创企业更关注通过采用务实的需求实践扩大客户群，从而推动发展。他们还以 Digital Port 生态系统中的十家初创企业为对象开展了案例研究。他们得出结论，即使初创企业随着客户群扩大而日趋成熟，仍会采用直接明了的需求工程方式。

Gralha et al. [25] 采用扎根理论方法，研究了 16 家软件初创企业的需求工程演变。他们发现，需求工程相关的演变涉及六个关键因素：需求制品、知识管理、需求相关角色、规划、技术债务和产品质量。此外，他们发现，一个维度的进步往往会促进其他维度的发展。不过，一个有趣的结论是，这些维度的演变并非初创企业成功的根本条件，但会对产品、员工和整个公司产生积极影响。

2) 设计：设计是软件开发的第二阶段，紧接着需求工程。在这一阶段，软件架构师和开发人员根据需求设计高层系统架构。一些软件初创企业遵循设计原则和方法论，而另一些可能不遵循。

初创企业采取以创始人为中心的方法，并且根据创始人的背景，他们可能会在开发阶段之前推动一些架构设计 [23]。由于创始人承担了风险，因此他们在决定设计方案及其他考量因素方面拥有更大的决定权。如果初创企业无法创造收入，就会面临巨大的后果。

Crowne 等人 [15] 发现，初创企业没有经验丰富的开发人员，因此会忽视架构和设计等非编码问题。初创企业还面临财务限制。因此，它们很难雇用经验丰富且高素质的人力资源，而这会直接影响软件的架构和设计。

Duc 等人 [19] 基于访谈、观察和文档记录，对五家早期初创企业开展了一项实证研究。他们发现，初创企业不了解最小可行产品（MVP）。最小可行产品有助于实现经济高效的产品设计，并弥合沟通鸿沟。

Deias 等人 [18] 发现，初创企业缺乏编写完善的架构与设计规范。这种缺乏部分是由于时间和资源限制。此外，初创企业面临尽快交付产品的压力，导致设计实践受到影响。

Souza 等人 [48]观察到，他们研究的所有初创企业都会与客户密切合作，并在一次快速会议中完成软件的简单设计。

Paternoster 等人 [41] 在其系统性研究中得出结论，使用知名框架有助于快速进行产品变更。此外，Jansen 等人 [30] 在对两家初创企业的研究中发现，对第三方软件进行机会主义且务实的复用有助于快速开发软件，从而缩短上市时间。此外，代码复用能够强化产品的架构结构，并提升产品的可扩展性。

3) 开发： 在软件开发的这一阶段，需求和设计会被落实为系统组件。开发人员编写代码并设计数据库，同时构建支持这些组件的 IT 基础设施。初创企业会采用各种开发实践来实现其产品。

Nicol`oPaternoster 等人 [41] 对初创企业的软件开发实践开展了一项系统性映射研究。他们发现，初创企业并不遵循任何标准的软件开发实践。这种倾向是合理的，因为初创企业首要关注的是尽可能早地将产品推向市场，以开始创收。此外，他们还希望将开发成本降至最低。因此，建立正式流程需要投入时间和资金。然而，这两者对于初创企业而言都是重要的限制因素。因此，初创企业通常采用不可预测、响应式且低精度的软件工程 [50] [31]。

Heitlager 等人 [26] 发现，初创企业通常具有一个共同模式：少数人员在资源匮乏的情况下起步。Coleman 等 [13]也观察到了同样的情况。初创企业资源有限，只将有限的资源用于支持产品开发。

Dande 等人 [17] 研究了芬兰和瑞典初创企业中的工作实践。他们发现，软件初创企业采用了 63 项常见实践。

Souza 等人 [48]通过对初创企业的首席技术官和首席执行官进行 14 次访谈，研究了软件初创企业中的敏捷开发。他们发现，工具和流程支持了大多数开发活动，以促进软件开发过程。例如，使用版本控制系统能够实现软件的持续集成和部署。

4) 测试： 测试用于检查软件是否实现预期功能，并确保整体质量。尽管质量保证在软件开发中至关重要，但大多数初创企业基本上都缺乏质量保证 [46]。

Pompermaier 等人 [42]对巴西科技园中的 8 家初创企业开展了一项研究。他们发现，这些初创企业的技术团队在第一个版本中没有进行任何测试。因此，测试取决于最终用户。但在后续版本中，75% 的初创企业采用了某种测试。

同样，Giardino 等人 [23] 发现，初创企业认为采用标准软件开发实践是在浪费时间。因此，他们忽略这些实践，以便尽快发布产品

知识领域	主要研究编号	摘要
需求工程	Chakraborty et al., 2012年; Melegati et al., 2016年; Rafiq et al., 2017年; Alves et al., 2020年; Gralha et al.,2018年;	初创企业通常没有任何正式流程; 一些初创企业可能有明确的需求工程步骤; 在需求实践较为成熟的情况下，这主要是由于市场压力或需求所致; 创始人在选择采用哪种实践方面发挥着至关重要的作用;
设计	Giardino 等人，2016年; Crowne 等人，2002年; Duc 等人，2016年; Deias 等人，2002年; Souza 等，2019年; Paternoster 等人，2014年; Jansen 等人，2008年;	初创企业采用以创始人为中心的方法 创始人承担了巨大的风险，因此在决定设计时拥有更大的话语权 初创企业缺乏经验丰富的开发人员，因此会忽视架构和设计等非编码问题
开发	Paternoster 等人，2014年; Sutton 等，2000年; Kakati 等，2003年; Heitlager 等人，2007年; Coleman 等，2008年; Dande 等，2014年; Souza 等人，2019年;	大多数初创企业不遵循任何标准的软件开发实践 创始人希望尽量降低开发成本。 要建立流程，需要投入时间和资金，而这两者都是初创企业面临的主要限制因素。 初创企业大多采用不可预测、响应迅速且精确度较低的软件工程方式
测试	Shikta 等，2021年; Pompermaier 等，2017年; Giardino 等人，2016年; Unterkalmsteiner 等人，2016年; Mater 等人，2000年; Shikta 等人，2021年; Thongsukh 等人，2017年;	初创企业通常忽视任何质量保证 最初几个版本的测试依赖于最终用户
部署	Silva 等人，2005年; Taipale 等人，2010年;	有些采用手动部署 有些采用 CI/CD 流水线

表 II
通用软件初创企业的软件工程实践总结

知识领域	主要研究编号	摘要
需求工程	E.d. S. Nascimento 等，2019年; A. Banks 等，2019年;	机器学习应用还有一些与训练数据相关的额外需求 由于客户无法恰当地理解 数据和度量需求，因此根据业务需求制定需求颇具挑战
数据管理	A.Hopkins 等人，2021年; A. Arpieg 等人，2018年;	准备训练数据时会出现各种数据缺陷，开发人员会使用各种 验证工具。 大多数初创企业难以实现数据标准化，因此会使用可信子集。
模型学习	A. Arpieg 等人，2018年; O. Simeone; D.Fox 等人，2021年;	初创企业在实验管理和深度学习模型故障排查方面面临挑战 他们还面临着在短时间内完成大量工作的压力，因此可能会使用多个 GPU 来训练模型
质量保证	L. E. Li 等人，2017年; E. d. S. Nascimento 等人，2019年;	初创企业倾向于将大型机器学习模型拆分为较小的模型，以避免训练期间出现各种缺陷 随着新数据不断加入，应定期监控模型
部署	L. E. Li 等人，2017年; A.Arpteg 等人，2018年 F. Ishikawa 等人，2019年;	机器学习模型对硬件的依赖性更高，而且不同位置可能需要不同的模型 由于初创企业的大多数客户并不倾向于采用模型版本管理，持续工程仍难以实现。

表 III 机器学习初创企业其他软件工程实践摘要

。因此，软件的首个版本中没有质量保证。Unterkalmsteiner 等人 [54] 也得出结论，初创企业通常会忽略任何质量保证活动。

Mater 等人 [37] 得出结论，如果组织内部缺乏资源，初创企业可以将质量保证测试外包给外部专家。这种外包方式可以有效地维持质量，也有助于降低成本并节省时间。

Shikta 等人 [46] 基于对孟加拉国初创企业的研究，提出了一个质量保证七阶段框架。各阶段为：(1) 引入质量保证负责人，负责确保需求和单元测试的质量；(2) 锁定需求并进行回归测试；(3) 聘用首位专职测试人员；(4) 实施测试度量指标；(5) 聘用第二位专职测试人员，负责处理来自客户的新需求和所需的缺陷修复；(6) 聘用第三位专职测试人员，负责实施自动化测试；(7) 制定定制测试计划。

Thongsukh 等人 [53] 指出，过程质量和产品质量与软件开发过程的质量密切相关。因此，采用 Scrum 方法等敏捷开发框架，有助于初创企业和大型企业提升产品质量。.

5) 部署： 软件部署包括将产品或交付物交付给最终用户所涉及的所有活动。这是产品开发的最后阶段。在这一阶段，产品已准备好在生产环境中使用。初创企业会采用各种部署模型。

Silva 等人 [16] 发现，一些软件初创企业会手动部署代码。另一些则使用持续集成工具进行部署，正如 Taipale 等人 [51] 所报告的那样。我们发现，关于部署实践的研究十分有限。

B. 研究问题*2*：机器学习初创公司还遵循哪些软件工程实践？

尽管通用软件初创企业与机器学习初创企业采用了相似的软件开发实践，但两者之间也存在一些差异。表 III 总结了机器学习初创企业遵循的其他实践。我们将在后续章节中对这些实践进行详细讨论。

1) 需求工程：首先，机器学习初创企业会尝试理解问题并定义目标 [39]。通用软件初创企业所采用的需求工程流程同样适用于机器学习初创企业。然而，机器学习应用还具有围绕训练数据的其他需求。Alec等人 [7] 提出了训练数据需求的九项额外考量。作者建议，训练数据应与高层需求相关，不应包含偏差，并且应当充分、自洽、可靠、稳健且正确。

然而，与通用软件初创企业类似，机器学习初创企业在根据业务需求创建需求时也面临挑战 [39]。问题主要在于客户无法正确理解为实现目标所需的指标和数据。这一挑战也源于用户的高期望 [27]。客户通常希望其产品的性能能够达到谷歌等大型科技公司的水平。但由于数据、计算资源和专业能力方面的限制，他们无法满足这些需求。

2）数据管理：数据管理是通用软件初创企业与机器学习初创企业之间的主要差异之一。对于机器学习公司而言，数据是项目的起点。Rob Ashmore等人 [6] 指出，数据管理阶段包含四项活动。第一项是数据收集，包括使用现有数据库或创建新数据库。第二项是预处理。在此阶段，会对收集到的数据进行适当调整。第三项是数据增强。在此阶段，会通过不同技术增加数据样本的数量。最后一步是分析收集到的数据和增强后的数据。开发人员可以构建处理结构化数据或非结构化数据的数据管道，以构建机器学习算法。在上述活动过程中可能会出现各种数据缺陷，开发人员可以使用数据验证工具来处理这些问题 [24]。

Aspen 等人 [28] 发现，几乎所有小型公司（包括初创企业）都难以规范数据录入并收集正确的数据，因此往往依赖一部分可信数据。然而，这些公司也发现了领域专家标注的数据存在问题。为了解决这一问题，其中大多数公司会制定复杂的流程并达成共识。此外，由于机器学习算法，尤其是神经网络，具有黑箱特性，机器学习公司在数据安全和隐私方面面临固有挑战 [4]。

3) 模型学习：机器学习初创企业的主要产品是训练后的机器学习模型。在模型学习阶段，会使用训练数据 [47]训练模型或机器学习算法。Rob 等人[6] 提出了该阶段的四项活动。第一项活动是从众多可用模型中选择合适的模型。第二项活动是训练，通过训练优化机器学习模型的性能。第三项活动是超参数选择，即在训练过程中选择参数。作者提出的第四项活动是迁移学习，即在多个领域中重复使用机器学习模型。

初创企业在这一阶段可能面临多项挑战。Anders 等人 [4]提出了深度学习应用在这一阶段面临的一些关键挑战。首要挑战是实验管理。作者建议在这一阶段跟踪硬件、平台、源代码配置、训练数据和模型状态。深度学习模型的故障排查也是一项主要挑战，因为在系统完成训练之前，很难估计其结果。此外，初创企业还面临着在短时间内取得大量进展的压力。Dylan 等人 [21] 建议快速训练和重新训练模型。初创企业可以使用更多图形处理器进行训练，或采用低精度训练。

4）质量保证：机器学习应用的主要目标之一是进行预测和推断，这使机器学习初创企业有别于一般初创企业 [36]。Rob 等人 [6] 提出了这一阶段的三项活动。第一项是需求编码，即将需求转换为可验证测试和数学性质。第二项是基于测试的验证，包括为经过训练的模型提供测试用例，并检查其是否输出预期结果。第三项是形式化验证，通过数学技术提供充分证据，证明模型满足需求。

在相同数据上多次训练机器学习模型时，可能会观察到不同的缺陷 [28]。Aspen 等人 [28]指出，谷歌和微软等大型公司通过多次训练模型来缓解这一问题，但小型公司和初创企业可能无力承担这样做所需的资源。在这种情况下，开发人员会将较大的模型分解为较小的模型。此外，训练好的模型在实时模式和批处理模式下应表现出相同的行为[34]。此外，在生产环境中，新的数据会持续插入机器学习模型，因此开发人员应继续监控模型性能 [39]。

5) 部署：Rob 等人 [6] 提出了机器学习应用部署阶段的三项活动。第一项是集成，即将机器学习模型集成到更广泛的系统架构中。第二项是监控，包括监控提供给模型的输入、监控环境、监控模型内部状态以及监控模型输出。最后，由于数据分布会随着时间推移而变化，因此应定期更新已部署模型。然而，对于许多初创企业而言，更新和模型版本管理仍然难以实现 [28]，因为持续工程化不完善的机器学习应用可能不受客户青睐 [29]。

与通用软件不同，机器学习应用对硬件的依赖程度更高，因为机器学习模型的性能取决于图形处理器 [5]。此外，各种模型需要部署在不同位置，以满足客户的需求 [34]。

V. 效度威胁

本节讨论了我们研究的局限性。

构念威胁：本研究仅考虑了2000年1月1日至2021年11月30日发表的论文。这一时间范围排除了此前和此后发表的所有论文，限制了我们的研究范围。此外，我们仅考虑了发表于期刊或会议论文集中的论文。由于关于机器学习初创企业的论文十分少见，我们还将一些相关博客纳入了研究。由于 IEEE Xplore 是我们的首选数据库，我们主要使用了该来源的论文，这可能造成了偏差。此外，我们忽略了新闻报道、视频、播客、访谈等所有灰色文献。不过，我们认为本研究的相关风险很小，因为大多数初创企业不会发表介绍其软件工程实践的论文。

外部威胁：本研究仅使用了37篇论文。尽管这一数量有助于我们深入了解一般软件初创企业和机器学习初创企业的软件工程实践，但样本量仍然较小。此外，由于专门提及初创企业的论文较少，尤其是关于机器学习初创企业的论文，我们采用了有关小型团队或公司的论文。小公司或小团队是初创企业的一个特征，但并不总能代表初创企业。因此，我们的研究结果可能无法推广到所有初创企业。

内部威胁：我们研究的主要内部有效性来自使用了一个论文池，该论文池通过特定关键词，并在 IEEE Xplore 和 ACM Digital Library 等可靠数据库中进行检索获得。随后，我们自动删除摘要中未包含全部关键词的论文，并通过阅读其余论文的摘要进行人工核验，从而进一步验证这些论文。我们还通过 Google Scholar 和 ml-ops.org 丰富了论文数据库。最后，我们还采用雪球抽样和人工检索，将相关论文加入我们的文献集合。因此，我们认为这一威胁很小。

VI. 讨论与结论

我们的研究首次对通用软件初创企业与机器学习初创公司之间的软件工程实践进行了比较研究。我们的目标是找出能够预测初创企业成功的软件工程实践。对于新的初创企业而言，遵循一种更严格且经过验证的指导方针将会有所帮助。通用软件初创企业也会更有信心地采用这些技术。但是，我们的研究表明，并不存在能够预测成功的流程，尤其是在公司的收入和存续时间方面。另一方面，我们发现，采用软件工程实践能够提高整体工作质量和员工满意度。从长远来看，采用适当的软件工程实践有助于更好地交付软件。

机器学习初创公司与一般初创企业采用相似的软件工程实践，同时也会遵循额外的实践来应对自身特有的挑战。主要差异源于机器学习应用所需的数据。这一差异使得在需求工程和数据管理阶段需要采取额外的实践。与一般初创企业所经历的开发阶段相对应，机器学习初创公司采用模型学习阶段。一般初创企业和机器学习初创公司都面临时间压力，需要在资源有限的情况下尽快交付各自的产品。因此，机器学习初创公司往往会将大型模型拆分成较小的模型，并使用多个 GPU 来赶上期限。

我们还旨在找出软件开发过程中各个阶段所使用的工具。遗憾的是，该领域研究的缺乏阻碍了这一目标的实现。在一些文献中，我们发现使用工具有助于改进软件开发过程。然而，我们未能找到足够的信息，无法为初创企业推荐一份详尽的工具清单。

未来，我们计划对选定的一般初创企业和机器学习初创公司开展案例研究，以了解它们在软件工程实践方面的差异，因为基于初创企业的论文数量有限。还可以考虑采访来自不同初创企业的开发人员，以深入了解这些初创企业如何在软件开发生命周期的各个阶段实施软件工程实践。此外，为弥补论文数量不足的问题，灰色文献（如对初创企业创始人的访谈、视频、播客以及 Medium 博客文章）可以成为了解初创企业当前采用的软件工程实践的优质资料来源。

参考文献

[1] B. Akter 和 M. A. Iqbal。平台型初创企业的失败因素：系统性文献综述。北欧媒体管理期刊，1(3):433–459，2020年。[2] C. Alves、J. Cunha 和 J. Ara´ujo。初创企业生态系统中需求工程实践的实用性探讨。载于*2020*年*IEEE*第*28*届国际需求工程会议（*RE*），第311–321页，2020年。[3] S. Amershi、A. Begel、C. Bird、R. DeLine、H. Gall、E. Kamar、N. Na-gappan、B. Nushi 和 T. Zimmermann。机器学习软件工程：一项案例研究。载于*2019*年*IEEE/ACM*第*41*届国际软件工程会议：软件工程实践（*ICSE-SEIP*），第291–300页。IEEE，2019年。[4] A. Arpteg、B. Brinne、L. Crnkovic-Friis 和 J. Bosch。深度学习的软件工程挑战。*2018*年*Euromicro*第*44*届软件工程与高级应用会议（*SEAA*），2018年8月。[5] A. Arpteg、B. Brinne、L. Crnkovic-Friis 和 J. Bosch。深度学习的软件工程挑战。载于 *2018* 年 *Euromicro* 第 *44* 届软件工程与高级应用会议（*SEAA*），第50–59页。IEEE，2018年。[6] R. Ashmore、R. Calinescu 和 C. Paterson。保障机器学习生命周期：目标、方法与挑战。*ACM*计算综述，54(5)，2021年5月。[7] A. Banks 和 R. Ashmore。机器学习中的需求保障。载于*SafeAI@ AAAI*，2019年。[8] L. Blake。影响人工智能生态系统与人工智能初创企业成功的因素：多层次分析。博士论文，HEC蒙特利尔，2020年。[9] T. Buganza、C. Dell'Era、E. Pellizzoni、D. Trabucchi 和 R. Verganti。揭示新技术所带来的潜力：在智能手机应用行业中实现技术顿悟的流程。创造力与创新管理，24(3):391–414，2015年。[10] M. Cantamessa、V. Gatteschi、G. Perboli 和 M. Rosano。初创企业的失败之路。可持续发展，10(7):2346，2018年。[11] A. Chakraborty、M. Baowaly、U. A Aref´ın 和 A. N. Bahar。需求工程在软件开发生命周期中的作用。《计算与信息科学新兴趋势期刊》，3 729，05 2012 723: – 。[12] M. Chui。人工智能：下一个数字前沿。麦肯锡全球研究院，47(3.6)，2017年。

[13] G. Coleman 和 R. V. O’Connor。对软件初创企业软件开发流程形成的调查。《企业信息管理杂志》，2008年。[14] M. Crowne。软件产品初创企业为何失败以及应对之道：初创公司软件产品开发的演变。载于IEEE国际工程管理会议，第1卷，第338–343页。IEEE，2002年。[15] M. Crowne。软件产品初创企业为何失败以及应对之道：初创公司软件产品开发的演变。载于IEEE国际工程管理会议，第1卷，第338–343页，第1卷。IEEE，2002年。[16] A. F. Da Silva、F. Kon 和 C. Torteli。赤道以南的XP：在巴西实施XP的经验。载于极限编程与软件工程敏捷过程国际会议，第10–18页。Springer，2005年。[17] A. Dande、V.-P. Eloranta、H. Hadaytullah、A.-J. Kovalainen、T. Lehtonen、M. Lepp¨anen、T. Salmimaa、M. Syeed、M. Vuori、C. Rubattel 等。软件初创企业模式：一项实证研究。2014年。[18] R. Deias、G. Mugheddu 和 O. Murru。在初创企业中引入XP。载于第三届极限编程与软件工程敏捷过程国际会议（XP）论文集，第62–65页，2002年。[19] A. N. Duc 和 P. Abrahamsson。最小可行产品还是多方面产品？最小可行产品在软件初创企业中的作用。载于敏捷软件开发国际会议。Springer，2016年。[20] V.-P. Eloranta。控制初创企业混乱的模式。载于第八届北欧程序模式语言会议（VikingPLoP）论文集，第1–8页，2014年。[21] D. Fox。初创企业如何训练大型深度学习模型，2021年。[22] M. Garbuio 和 N. Lin。人工智能作为医疗保健初创企业的增长引擎：新兴商业模式。《加利福尼亚管理评论》，61(2):59–83，2019年。[23] C. Giardino、N. Paternoster、M. Unterkalmsteiner、T. Gorschek 和 P. Abrahamsson。初创公司的软件开发：绿地初创企业模式。《IEEE软件工程汇刊》，42(6):585–604，2016年。[24] G. Giray。从软件工程视角看机器学习系统工程：最新进展与挑战。《系统与软件杂志》，第111031页，2021年。[25] C. Gralha、D. Damian、A. Wasserman、M. Goul˜ao 和 J. Ara´ujo。软件初创企业需求实践的演变。载于2018年IEEE/ACM第40届国际软件工程会议（ICSE），第823–833页，2018年。[26] I. Heitlager、R. Helms 和 S. Brinkkemper。研究和规划产品协同演化的一种试探性技术。载于第三届IEEE软件演化研讨会，2007年，第42–47页，2007年。[27] A. Hopkins 和 S. Booth。大型科技公司之外的机器学习实践：资源限制如何挑战负责任的开发。载于2021年AAAI/ACM人工智能、伦理与社会会议论文集，AIES ’21，第134–145页，美国纽约州纽约，2021年。计算机协会。[28] A. Hopkins 和 S. Booth。大型科技公司之外的机器学习实践：资源限制如何挑战负责任的开发。2021年。[29] F. Ishikawa 和 N. Yoshioka。工程师如何看待机器学习系统工程中的困难？——问卷调查。载于2019年IEEE/ACM第七届产业实证研究联合研讨会（CESI）和第六届软件工程研究与产业实践研讨会（SER IP），第2–9页，2019年。[30] S. Jansen、S. Brinkkemper、I. Hunink 和 C. Demir。创新型初创公司中的务实、机会主义复用。《IEEE软件》，25(6):42–49，2008年。[31] M. Kakati。高科技新创企业的成功标准。《技术创新》，23(5):447–457，2003年。[32] G. Kalyanasundaram。初创企业为何失败？班加罗尔的一项案例研究实证分析。《亚洲创新与政策杂志》，2018年。[33] E. Klotins、M. Unterkalmsteiner、P. Chatzipetrou、T. Gorschek、R. Prik-ladnicki、N. Tripathi 和 L. Pompermaier。初创企业技术债务的探索。载于2018年IEEE/ACM第40届国际软件工程会议：软件工程实践专题（ICSE-SEIP），第75–84页，2018年。[34] L. E. Li、E. Chen、J. Hermann、P. Zhang 和 L. Wang。扩展机器学习即服务。载于国际预测应用与API会议，第14–29页。PMLR，2017年。[35] L. Luce。时尚人工智能：人工智能如何革新时尚产业。Apress，2018年。[36] S. Masuda、K. Ono、T. Yasue 和 N. Hosokawa。《机器学习应用的软件质量调查》。载于《2018年IEEE国际软件测试、验证与确认研讨会（ICSTW）》，第279–284页，2018年。[37] J. L. Mater 和 B. Subramanian。《解决互联网初创企业的软件质量管理问题》。主题演讲，2000年10月17日。[38] J. Melegati 和 A. Goldman。《软件初创企业中的需求工程：一种扎根理论方法》。载于《2016年国际工程、技术与创新会议/IEEE国际技术管理会议（ICE/ITMC）》，第1–7页，2016年。[39] E. d. S. Nascimento、I. Ahmed、E. Oliveira、M. P. Palheta、I. Steinmacher 和 T. Conte。《理解机器学习系统的开发流程：挑战与解决方案》。载于《2019年ACM/IEEE国际实证软件工程与度量研讨会（ESEM）》，第1–6页，2019年。[40] C. Nobel。《公司为何失败，以及创始人如何东山再起》。哈佛商学院，马萨诸塞州波士顿，2011年。[41] N. Paternoster、C. Giardino、M. Unterkalmsteiner、T. Gorschek 和 P. Abrahamsson。《初创公司中的软件开发：一项系统性映射研究》。《信息与软件技术》，2014年。[42] L. Pompermaier、R. Chanin、A. H. C. de Sales、K. Fraga 和 R. Prikladnicki。《关于软件工程与软件初创企业的实证研究：来自创新生态系统案例的发现》。org. crossref. xschema. 1. Title@ 1ff57c40，2017年，巴西，2017年。[43] U. Rafiq、S. S. Bajwa、X. Wang 和 I. Lunesu¯。《应用于软件初创企业的需求获取技术》。载于《2017年第43届Euromicro软件工程与高级应用会议（SEAA）》，第141–144页，2017年。[44] M. Schulte-Althoff、D. F¨urstenau 和 G. M. Lee。《人工智能初创企业的规模化视角》。载于《第54届夏威夷国际系统科学会议论文集》，第6515页，2021年。[45] R. Shams。《在初创企业中更好、更快地开发机器学习产品》。《IEEE工程管理评论》，46(3)：36–39，2018年。[46] S. Shikta、H. M. Mahir Shahriyar、S. K. Das、S. N. Mahal、K. B. Al Jannat 和 S. Alam。《成功初创企业的质量保证指南》。载于《2021年IEEE/ACIS第19届软件工程研究、管理与应用国际会议（SERA）》，第81–85页，2021年。[47] O. Simeone。《机器学习及其在通信系统中的应用简介》。《IEEE认知通信与网络汇刊》，4(4)：648–664，2018年。[48] R. Souza、L. Rocha、F. Silva 和 I. Machado。《调查软件初创企业中的敏捷实践》。载于《第33届巴西软件工程研讨会论文集，SBES 2019》，第317–321页，纽约州纽约，美国，2019年。计算机协会。[49] J.-C. Spender、V. Corvello、M. Grimaldi 和 P. Rippa。《初创企业与开放式创新：文献综述》。《欧洲创新管理期刊》，2017年。[50] S. Sutton。《流程在软件初创企业中的作用》。《IEEE软件》，17(4)：33–39，2000年。[51] M. Taipale。《Huitale——一家芬兰精益初创企业的故事》。载于《国际精益企业软件与系统会议》，第111–114页。Springer，2010年。[52] D. A. Tamburri。《可持续机器学习运维：趋势与挑战》。载于《2020年第22届科学计算符号与数值算法国际研讨会（SYNASC）》，第17–23页，2020年。[53] S. Thongsukh、S. D. N. Ayuthaya 和 S. kiattisin。《基于Scrum框架的初创企业框架》。载于《2017年数字艺术、媒体与技术国际会议（ICDAMT）》，第458–463页，2017年。[54] M. Unterkalmsteiner、P. Abrahamsson、X. Wang、A. Nguyen Duc、S. Shah、S. Sohaib、G. Baltes、K. Conboy、E. Cullina、D. Dennehy、H. Edison、C. Fern´andez、J. Garbajosa、T. Gorschek、E. Klotins、L. Hokkanen、F. Kon、M. I. Lunesu、M. Marchesi 和 A. Yag¨ue。《软件初创企业——研究议程》。《E-Informatica软件工程期刊》，2016：89–124，2016年10月。[55] M. Unterkalmsteiner、P. Abrahamsson、X. Wang、A. Nguyen-Duc、S. Q. Shah、S. S. Bajwa、G. H. Baltes、K. Conboy、E. Cullina、D. Dennehy 等。《软件初创企业——研究议程》。《E-Informatica软件工程期刊》，第10卷第1期：89–123，2016年。[56] C. Vijai 和 W. Wisetsri。《印度医疗保健初创企业中人工智能的崛起》。《管理进展》，14(1)：48–52，2021年。[57] X. P. S. Zhang 和 D. Kedmey。《萌芽的浪漫：金融与人工智能》。《IEEE多媒体》，25(4)：79–83，2018年。