SEMODS：经过验证的开源软件工程模型数据集

Alexandra González、Xavier Franch、Silverio Martínez-Fernández 加泰罗尼亚理工大学，西班牙巴塞罗那
{alexandra.gonzalez.alvarez,xavier.franch,silverio.martinez}@upc.edu

arXiv:2601.00635v1 [cs.SE] 2 Jan 2026

摘要

将人工智能融入软件工程需要收集经过筛选、适用于软件工程任务的模型。Hugging Face（HF）上托管着数百万个模型，且新模型不断涌现；若没有专门的目录，便无法切实可行地识别软件工程模型。为弥补这一空缺，我们推出 SEMODS：一个聚焦软件工程的数据集，包含从 HF 提取的 3,427 个模型。该数据集通过自动化收集，并结合人工标注和大语言模型辅助进行严格验证。我们的数据集将模型与软件开发生命周期中的软件工程任务和活动关联起来，以标准化方式呈现模型评估结果，并支持数据分析、模型发现、基准测试和模型适配等多种应用。

关键词

面向软件工程的人工智能、模型、软件开发生命周期、Hugging Face

ACM 引用格式：

Alexandra González、Xavier Franch、Silverio Martínez-Fernández。2026 年。SE-MODS：经过验证的开源软件工程模型数据集。载于第三届 *ACM 人工智能基础模型与软件工程国际会议论文集（FORGE 2026）*。ACM，纽约，纽约州，美国，5 页。https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 引言

预训练模型（PTM）是指在特定数据集上，使用定义明确的数据处理和训练流水线进行训练的深度神经网络架构，由此获得经过学习的模型参数（权重） [11]。在软件项目中，预训练模型可作为核心组件，也可用于实验 [25]。

预训练模型通过机器学习（机器学习）注册库共享，这些注册库也称为模型中心或模型动物园，团队可在其中协作并共享机器学习资产 [22, 24]。这些注册库在促进复用方面发挥着关键作用，因为它们降低了从头训练模型所需的成本和精力，同时促进了可复现性 [11]。

Hugging Face (HF) [18]是一个广受欢迎的开源注册平台，旨在分享和开发模型、数据集以及基于它们构建的应用（称为 Spaces）。HF 中的每个模型仓库都存储着丰富的属性，涵盖流行度指标（例如点赞数、下载量）以及许可证、库、训练数据集和推理提供商等元数据。凭借其开放性和活跃的社区，该平台实现了显著增长。根据我们自行计算的结果，仅在 2024 年，HF 每天平均新增 2,199 个模型，每日下载量超过六百万次。截至 2025年11月，该平台托管的模型已超过两百万个 [26]，并且仍在快速增长，平均每十五秒就会创建一个新仓库 [12]。

尽管拥有丰富的资源，软件工程研究人员 [14] 和从业者 [4, 35, 36] 仍难以找到与其任务直接相关、同时满足项目约束（如许可或性能要求）的模型。常见问题包括属性缺失、报告性能与实际性能不符，以及与隐私或不道德模型行为相关的风险 [22]。机器学习模型注册库 [13] 缺乏面向软件工程的目录，限制了机器学习融入软件开发生命周期 (SDLC) [32], ，迫使用户手动浏览海量模型集合以寻找合适的候选模型。这一过程既耗时又容易出错，最终拖慢了机器学习在软件工程工作流中的应用 [5]。

为弥补这一空白，我们推出软件工程模型（SEMODS）：一个软件工程模型数据集，经过系统收集、处理和验证，旨在支持对这些模型进行深入分析、高效发现和在软件工程场景中的实际应用。该数据集使研究人员和从业者能够探索针对特定软件工程任务的模型，并通过 SQL 查询与相关元数据交互，同时支持基准测试，并促进模型适配所需模型的识别。

本研究的贡献如下。
(1) 一个经验证的数据集，包含 3,427 个 HF 软件工程模型，并按照软件开发生命周期中的软件工程活动和任务进行编目。(2) 对基准和指标进行标准化表示。(3) 自动化相关流程，使数据集能够及时纳入新的 HF 仓库并更新动态属性。

数据可用性：数据集的快照（2025年11月版）包含 3,427 个软件工程模型及其相关属性，可在 Zenodo [7] 中公开获取。为促进可复现性，完整的编目流水线也可在 Zenodo [6] 中获取。

2 相关工作

已有多项工作收集并整理了模型注册库中的数据，以便重复利用模型并开展大规模分析，其中包括与软件工程相关的模型。

Ait 等人 [2] 介绍了 HFCommunity，这是一个从 HF Hub 汇总数据的关系数据库。该数据库旨在解决缺少可在平台 API 之外收集和探索 HF 数据的工具这一问题。

允许免费为个人或课堂使用而制作本作品的全部或部分数字副本或纸质副本，前提是这些副本不得用于或分发以获取利润或商业利益，并且副本须附有本声明及首页上的完整引文。必须尊重本作品中由作者以外的其他人拥有版权的部分。注明出处后，允许摘录。若要以其他方式复制、再版、发布到服务器或分发给列表中的对象，须事先获得特定许可和/或支付费用。请向 permissions@acm.org 申请许可。
FORGE 2026，巴西里约热内卢© 2026 版权所有归所有者/作者所有。出版权已授权给 ACM。
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

同样，Jiang 等人 [23] 提出了 PTMTorrent，这是一个旨在促进预训练模型软件包评估与理解的数据集，涵盖预训练权重、文档、模型架构、数据集和元数据。PTMTorrent 整合了五个模型中心（HF、Model Zoo、PyTorch Hub、ONNX Model Zoo）的信息，覆盖 15,913 个软件包。由于篇幅所限，HF 子集仅包含下载量最高的 10% 模型。

随后，Jiang 等人 [24] 发布了 PeaTMOSS 数据集，其中包含 281,638 个预训练模型的元数据，以及月下载量超过 50 次的模型的详细快照（14,296 个预训练模型）。PeaTMOSS 还将预训练模型与 GitHub 上使用这些模型的 28,575 个开源软件仓库关联起来，在 15,129 个下游 GitHub 仓库与相应的 2,530 个预训练模型之间建立了 44,337 个映射。这种模型与仓库数据的整合，为挖掘预训练模型及研究预训练模型供应链提供了机会。

与以往工作相比，我们专门筛选并验证了与软件工程明确相关的模型，提出了与软件工程任务的新映射，并提供了标准化的基准测试数据。我们的数据集为用户提供了一种高效发现和获取与其任务相关的软件工程模型的方式。这项工作扩展了现有数据集，例如 HFCommunity [2]和 PeaTMOSS [24]，纳入了聚焦软件工程的目录，并对评估信息进行了协调，从而支持在软件开发生命周期中有针对性地探索和复用预训练模型。

3 数据集构建与应用

图1展示了用于构建数据集的流程，并说明了其应用。下文将详细介绍编目流程，包括任务识别、数据收集、处理和验证；随后通过概念模式描述数据集的内容。接下来，我们将介绍如何通过监测新模型并更新动态属性来维护数据集，使其保持最新，并讨论其主要应用。编目流水线最初基于截至2025年3月的 Hugging Face 模型完整快照运行，涵盖150万个模型；经过验证后，它成为一套自动化工作流，每天对新发布的资源执行相同的步骤。

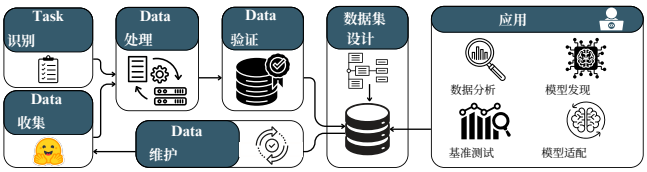

图1：数据集构建流程与应用。

3.1 任务识别

我们在既有研究的基础上，构建了涵盖软件开发生命周期五个阶段的软件工程任务分类体系。从 Hou 等人识别出的 88 项任务入手 [17]，，我们将其与 Sommerville 等既有来源对齐 [32]，，其中包括将“软件开发活动”更名为“软件实现”。我们进一步完善了分类体系，将软件架构视为软件设计的一部分，并根据《软件工程知识体系》的章节添加任务 [16]。最终形成的分类体系，经由作者讨论以确保覆盖全面且避免重叠后，包含 147 项任务。我们注意到，目前的 Hugging Face 模型覆盖了 100 项软件工程任务，但由于编目流水线每天运行，随着新模型发布，覆盖范围可能会逐渐扩大。

3.2 数据收集

我们通过其 API [21] 获取了托管在 Hugging Face 上的所有模型。对于每项资源，我们都考察了与之相关的所有可用文档。具体而言，在以下内容存在时，我们均予以收集：(i) 其模型卡描述，即一份用于说明模型特征、预期用途和评估情况的 Markdown 文件 [27]；(ii) 相关的卡片元数据，以 YAML 代码块指定（例如许可证、标签、语言） [20]；以及 (iii) 链接的 arXiv 论文的摘要 [19]，，并利用了 Hugging Face 与 arXiv 之间的跨平台链接 [33]。

3.3 数据处理

收集数据后，我们对其进行了处理，以便自动编目。首先，我们进行了文本规范化，包括分词、小写化和词形还原，以便准确检测模型文档中的软件工程任务。由于我们聚焦于与软件工程相关的资源，因此在处理后的文本中搜索软件工程任务：对于多词任务（例如“代码生成”），要求所有词语完整且按正确顺序出现；对于仅作为较长词元一部分出现的部分匹配，则予以排除。为确保匹配结果严谨可靠，我们识别了异常值（即出现频率异常高的软件工程任务），并找出因模型文档条目之间文本相似度较高而产生的重复实例，从而避免重复记录或对同一预训练模型进行多次计数。

3.4 数据验证

最后，我们进行了严格验证，以确保仅保留与软件工程相关的资源；验证分为两个阶段，结合了人工标注和大语言模型（LLM）辅助。第一阶段，我们针对每项软件工程活动随机抽取一部分模型，并确保涵盖所有相关任务。我们采用 95% 置信水平和 5% 误差范围计算样本量，以确保统计有效性 [30]。第一作者对所有这些子集进行了人工标注，共计 1,346 个模型。根据定性研究中既有的标注者间信度实践 [28]，，另外两位具有领域经验的作者对每项软件工程活动中 10% 的样本进行了标注。由此得到的标注数据作为大语言模型（Gemini 2.0 Flash）的真实标准；我们在零样本设置下对其进行提示，要求其对试点集中的每个预训练模型给出二元相关性判断和判断依据。在使用科恩 κ 系数评估模型性能 [31]，并不断完善全部五个提示，直至获得近乎完美的一致性水平（$k > 0.8$）后，第二阶段验证在此前未见过的数据上测试了大语言模型的泛化能力。测试确认模型可靠后，我们使用大语言模型判断完整数据集中每个模型是否涉及某项软件工程活动。

3.5 数据集设计

数据集模式定义为 UML 类图，如图 2 所示。该表示从软件工程（SE）的视角对 Hugging Face 仓库进行结构化，使其能够支持查询，并促进对软件工程中机器学习生态系统的分析和有效利用。



图 2：数据集的概念模式。该设计引入了软件工程模型特有的新实体（例如 SETask）及其标准化评估表（例如基准测试、指标），从而支持软件工程模型的开发与评估。

通过 Repository 类及其子类 Model，层次结构表示仓库类型，其中 Model 专门表示模型。为支持未来扩展数据集，使其包含模型之外的其他 HF 类别（例如数据集或 Spaces），Repository 类包含属性 repository_type，用于指定 HF 中的仓库类别。每个仓库实例都存储可识别的元数据（我们的内部 repository_id 以及由 HF 定义的外部_id），以及描述 HF 仓库特征的属性。这些属性包括时间信息（created_at）、受欢迎度指标（下载量和点赞数的原始值与归一化值，以及由此推导出的受欢迎度）和活动度量（num_commits 和 num_contributors）。Model 子类通过增加 arxiv（用于表示论文标识符）和大小等属性，扩展了这些信息。由于该平台不会直接提供模型大小，我们根据与每个模型关联的文件总大小，以字节为单位计算模型大小。

仓库与四个用于描述其特征的实体相关联。第一个是 SETask 类，它指定预训练模型所支持的软件工程活动（activity）和任务（name）。activity 属性的可能值是封闭的，并且与软件开发生命周期的五个阶段（SEActivityType）直接对应。此外，单个模型不一定局限于某一项特定的软件工程任务或活动，因为许多模型在整个生命周期中都具有广泛的用途。当前 HF 元数据中缺少这种 HF 仓库与特定软件工程活动和任务之间的映射；该映射是通过数据集构建前序步骤中详述的编目流程得出的。这些信息对于将机器学习资产集成到软件工程中尤其有用，因为它将模型的用途直接映射到软件开发生命周期的特定环节。最后，RepositorySETaskLink 类正式关联仓库与检测到的软件工程任务，并包含一个 reasoning 属性，用于描述从模型文档中提取的证据。

第二个描述性实体是 Benchmark 类，它通过 RepositoryBenchmarkLink 类包含每个仓库的结构化评估信息。尽管模型卡会在模型索引部分提供结果，但这些信息存在异质性，且缺少标准化报告。为解决这一问题，我们对 Hugging Face 文档中不一致的命名应用了基于规则的规范化（例如，统一基准测试的不同名称，如 大规模多任务语言理解 和 大规模多任务语言理解，或指标名称，如 准确率_归一化准确率 和 归一化_准确率）。对于每个具有可用评估数据的仓库，我们获取了基准测试_名称、实现框架、编程语言、性能指标、其配置（config）以及对应的数值得分。我们收集到 206 个不同的基准（例如，HumanEval [10], MBPP [8]）和 43 种不同的指标（例如，归一化准确率、余弦准确率）。

第三个实体是标签类，这是 Hugging Face 中的一项关键功能，可提供与预训练模型 [34] 相关的附加信息。每个标签实例都指定其类别（category_name）、描述性标签、类型（tag_type）和对应的值。标签类别涵盖多种存储库属性，包括训练数据集、可用的推理_提供方、支持的语言、底层库、许可证、关联的流水线_标签、部署区域和大小_类别等。标签通过存储库标签关联类与仓库关联，因此每个仓库都可以关联多个标签。

最后，我们介绍提交类。作为第四个描述性实体，它记录对仓库所做更改的信息。提交类包含一个唯一标识符（commit_id）、时间信息（created_at），以及用于存储提交详情的文本字段，包括

标题和消息等。仓库通过仓库提交关联类与这些记录关联，该类还包含一个作者属性，用于记录更改的执行者。鉴于模型经常被重复使用或分叉，同一个提交可以通过此关联表与多个仓库相关联。

利用收集、处理和验证后的数据以及概念模式，我们生成了 SEMODS 表。两张辅助表支持这一过程：一张存储从 HF API 检索到的原始仓库，另一张仅包含软件工程仓库。

3.6 数据维护

鉴于 Hugging Face 生态系统的快速发展 [3, 15] 以及新仓库和模型贡献不断增加 [9], , 我们实施了一套自动化流程，每天检查是否有新发布的 Hugging Face 模型。每个新仓库都会自动编目；若被判定与软件工程无关，则会被舍弃。此外，动态属性（例如点赞数、下载量、num_次提交、num_位贡献者）每天更新两次，以确保指标准确。

3.7 应用

SEMODS 的关系结构和对软件工程（SE）的特定关注，为从事软件工程人工智能（AI）研究的研究人员和从业者提供了多种机会。这些数据能够支持新的实证研究，并为将人工智能融入软件开发生命周期提供实际应用场景。下文概述其主要应用，并列举可利用这些数据探讨的研究问题（RQ）。

3.7.1 数据分析。 SEMODS 支持对 HF 中的软件工程模型进行定量和定性分析，从软件工程视角刻画机器学习生态系统。定量分析可以揭示模型创建和复用方面的趋势，而定性探索则可以考察仓库文档或基准测试实践。结构化评估数据的可用性还能带来一些洞见，例如模型性能、大小与受欢迎度之间的相关性。将 SEMODS 与通用数据集进行交叉比对，凸显了其专业化范围：我们的软件工程模型中，只有 244 个（7.12%）和 990 个（28.89 %）分别与 PeaTMOSS [24]和 HFCommunity [2], 重叠。这种极低的重叠比例（占这些数据集的 <0.2%）源于 SEMODS 的时效性（PeaTMOSS 自 2023 年 8 月以来未更新，HFCommunity 自 2024 年 10 月以来未更新）和包容性（无论受欢迎度如何，SEMODS 都保留专业化和新兴模型）。这些分析有助于探讨以下研究问题：谁在 *HF* 上创建软件工程模型（例如，新手还是经验丰富的开发者）？模型维护状况与软件工程模型的持续受欢迎度有何关系？哪些基准用于评估软件工程模型？

3.7.2 模型发现。用户可以使用软件工程特定属性（例如软件工程任务、软件工程活动）查询 SEMODS，从而找到可集成到软件工程流程中的模型。例如，用户可以搜索支持“软件设计”或“代码摘要”任务的模型，同时要求其符合开放政策标准或指标限制。纳入推理属性有助于提高透明度，说明每个预训练模型如何映射到软件工程活动，从而支持基于语义、指标和学习的方法进行选择 [37], , 并减少大规模人工探索。由此可以提出以下问题：现有模型对哪些软件工程活动的支持最多、最少？按特定属性筛选时，模型发现模式会如何变化？

3.7.3 基准测试。通过提供涵盖异构基准和指标的结构化评估结果，用户可以针对预训练模型的性能开展实证研究。研究人员可以进行跨基准比较，以探究性能差异，也可以比较给定基准和配置下的结果，从而找出表现最佳的模型。此外，这些数据还可以揭示哪些基准常用于特定的软件工程活动和任务，并指出尚未充分探索、可能需要新基准的领域。这些能力有助于回答以下研究问题：模型在不同基准上的结果一致性如何？哪些软件工程活动缺乏系统化评估资源，或需要新的评估资源？

3.7.4 模型适配。 由于每个预训练模型都关联着多个属性，该数据集有助于通过微调、迁移学习 [38], 和知识蒸馏 [11, 37]等方法，识别可供适配的候选模型。用户可以查找在相似数据集或软件工程任务上训练的模型，为开发软件工程专用模型提供起点。例如，若开发者希望微调模型以进行代码编辑 ，便可查看现有代码生成模型的元数据，从而选出合适的候选模型，减少训练时间、计算成本和能源消耗。这些属性为以下研究问题提供了契机：哪些模型最适合适配到特定的软件工程任务？训练数据集的相似性会如何影响适配结果？

4 效度威胁

我们承认 SEMODS 存在一些效度威胁。对于文档不完善或缺失的模型，内部效度可能受到影响；而模型卡完整的模型则有充分的文档支持。我们通过在数据收集过程中补充外部证据（例如链接的 arXiv 摘要）来降低这一风险。外部效度仅限于 HF Hub，并可能受到 HF 未来文档实践变化的影响。构念效度取决于我们采用的软件工程任务分类体系，该体系以既有文献为基础 [16, 17]。最后，关于结论效度，我们设计了由三名独立人工标注员参与的验证协议（见第 3.4 节），以应对使用大语言模型确认软件工程相关性所带来的风险。

5 结论与未来工作

我们介绍了 SEMODS，这是一个包含 Hugging Face 中 3,427 个软件工程模型信息的数据集。数据通过严格的流程收集，该流程扫描平台上所有可用的模型，并依据软件开发生命周期对其进行分类。除 Hugging Face 已提供的元数据外，我们的数据集还新增了模型与特定软件工程任务和活动之间的映射，并通过提取和统一基准测试及指标配置，为报告的评估指标提供了标准化表示。

未来，我们计划通过增加其他信息来源（例如 GitHub [1], PyTorch Hub [29]）来扩展数据集。我们还计划将该数据集与提供互补属性的现有数据集相结合，并开发一个推荐系统，帮助用户识别可能感兴趣的模型。

6 致谢

本研究获得由西班牙大学、研究与创新部/西班牙国家研究署/10.13039/501100011033 资助的项目 PID2024-156019OB-I00 以及欧盟欧洲区域发展基金的支持。Alexandra González 还感谢 FI-STEP 项目资助 2025 STEP-00407。

参考文献

[1] GitHub Models - GitHub 文档 — docs.github.com。https://docs.github.com/en/ github-models。 [访问日期：03-11-2025]。[2] Ait, A., Izqierdo, J. L. C., 和 Cabot, J. HFCommunity：用于分析 Hugging Face Hub 社区的工具。载于 *2023*年 *IEEE* 软件分析、演化与再工程国际会议（*SANER*）（2023），第 728–732 页。[3] Ait, A., Izqierdo, J. L. C., 和 Cabot, J. HFCommunity：用于分析 Hugging Face Hub 数据的提取流程和关系数据库。计算机编程科学 *234*（2024），103079。[4] Ajibode, A., Bangash, A. A., Cogo, F. R., Adams, B., 和 Hassan, A. E. 探索 Hugging Face 上开放式预训练语言模型发布的语义化版本。实证软件工程 *30*，3（2025），1–63。[5] Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., 和 Zimmermann, T. 机器学习的软件工程：一个案例研究。载于 *2019*年 *IEEE/ACM* 第*41*届软件工程国际会议：软件工程实践（*ICSE-SEIP*）（2019），第 291–300 页。[6] 匿名作者。《SEMODS：经过验证的开源软件工程模型数据集》复现包。https://zenodo.org/records/17674909，2025年11月。[7] 匿名作者。SEMODS：经过验证的开源软件工程模型数据集（2025年11月快照）— zenodo.org。https://zenodo.org/records/ 17675256，2025年11月。[8] Austin, J., Odena, A., Nye, M., Bosma, M., Michalewski, H., Dohan, D., Jiang, E., Cai, C., Terry, M., Le, Q., 等。利用大语言模型进行程序合成。*arXiv* 预印本 *arXiv:2108.07732*（2021）。[9] Castaño, J., Martínez-Fernández, S., Franch, X., 和 Bogner, J. 分析 Hugging Face 上机器学习模型的演化与维护。载于 第*21*届国际软件仓库挖掘会议论文集（美国纽约州纽约市，2024），MSR '24，计算机协会，第 607–618 页。[10] Chen, M. 评估基于代码训练的大语言模型。*arXiv* 预印本 *arXiv:2107.03374*（2021）。[11] Davis, J. C., Jajal, P., Jiang, W., Schorlemmer, T. R., Synovic, N., 和 Thiruvathukal, G. K. 重用深度学习模型：软件工程中的挑战与发展方向。载于 *2023*年 *IEEE John Vincent Atanasoff* 现代计算国际研讨会（*JVA*）（2023），IEEE，第 17–30 页。[12] Delangue, C. Hugging Face 上的公开模型数量刚刚突破 1,500,000 — linkedin.com。https://huggingface.co/posts/clem/238420842235482/。 [访问日期：29-05-2025]。[13] Di Sipio, C., Rubei, R., Di Rocco, J., Di Ruscio, D., 和 Nguyen, P. T. 软件工程中预训练模型的自动分类：基于 Hugging Face 数据集的案例研究。载于 第*28*届软件工程评估与评价国际会议论文集（美国纽约州纽约市，2024），EASE '24，计算机协会，第 351–356 页。[14] Giray, G. 工程机器学习系统的软件工程视角：现状与挑战。*JSS 180*（2021），111031。[15] Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., 等。预训练模型：过去、现在与未来。*AI Open 2*（2021），225–250。[16] Hiroyasu Washizaki，编。软件工程知识体系指南（*SWEBOK* 指南），第*4.0*版。IEEE计算机学会，2024。[17] Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J., 和 Wang, H. 面向软件工程的大语言模型：系统性文献综述。*ACM* 软件工程与方法学汇刊（2024年9月）。已接受，即将发表。[18] Hugging Face。Hugging Face——建设未来的 AI 社区。— huggingface.co。https://huggingface.co， [无日期。]。 [访问日期：30-10-2025]。[19] Hugging Face 公司。模型卡——关联论文 — huggingface.co。https://huggingface.co/docs/hub/model-cards#linking-a-paper。 [访问日期：30-10-2025]。[20] Hugging Face 公司。模型卡——模型卡片元数据 — huggingface.co。https://huggingface.co/docs/hub/model-cards#model-card-metadata。 [访问日期：30-10-2025]。[21] Hugging Face 公司。Hugging Face Hub 文档 — huggingface.co。https: //huggingface.co/docs/hub/index， [无日期。]。 [访问日期：30-10-2025]。[22] Jiang, W., Synovic, N., Hyatt, M., Schorlemmer, T. R., Sethi, R., Lu, Y.-H., Thiruvathukal, G. K. 和 Davis, J. C. 对 Hugging Face 深度学习模型注册表中预训练模型复用情况的实证研究。载于《2023年 IEEE/ACM 第45届国际软件工程会议（ICSE）论文集》（2023），IEEE，第2463–2475页。[23] Jiang, W., Synovic, N., Jajal, P., Schorlemmer, T. R., Tewari, A., Pareek, B., Thiruvathukal, G. K. 和 Davis, J. C. PTMTorrent：用于挖掘开源预训练模型包的数据集。载于《 2023 年 IEEE/ACM 第20届国际软件仓库挖掘会议（MSR）论文集》（2023），IEEE，第57–61页。[24] Jiang, W., Yasmin, J., Jones, J., Synovic, N., Kuo, J., Bielanski, N., Tian, Y., Thiruvathukal, G. K. 和 Davis, J. C. PeaTMOSS：开源软件中预训练模型的数据集及初步分析。载于《第21届国际软件仓库挖掘会议论文集》（2024），第431–443页。[25] Koohjani, M. 和 Costa, D. E. 探索开源软件仓库中预训练模型的生命周期与维护实践。arXiv 预印本 arXiv:2504.06040（2025）。[26] Laufer, B., Oderinwale, H. 和 Kleinberg, J. 机器学习生态系统剖析：Hugging Face 上的200万个模型。arXiv 预印本 arXiv:2508.06811（2025）。[27] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D. 和 Gebru, T. 用于模型报告的模型卡片。载于《公平性、问责制与透明度会议论文集》（2019），第220–229页。[28] O'Connor, C. 和 Joffe, H. 定性研究中的编码者间信度：争论与实用指南。《国际定性方法杂志》19（2020），1609406919899220。[29] PyTorch 基金会。PyTorch Hub — pytorch.org。https://pytorch.org/hub/， [无日期。]。 [访问日期：03-11-2025]。[30] Qualtrics。样本量计算器 - Qualtrics — qualtrics.com。https://www.qualtrics.com/blog/calculating-sample-size/。 [访问日期：21-05-2025]。[31] Sim, J. 和 Wright, C. C. 信度研究中的 κ 统计量：用途、解释与样本量要求。《物理治疗》85卷3期（2005），257–268。[32] Sommerville, I.《软件工程》，第10版。Pearson，2015。[33] Suryani, M. A., Karmakar, S. 和 Mathiak, B. 通过异构信息网络探索 Hugging Face 模型，并在学术仓库间建立关联。载于《社会网络分析与挖掘进展国际会议论文集》（2024），Springer，第371–386页。[34] Suryani, M. A., Karmakar, S., Mathiak, B. 和 Mayr, P. 从 Hugging Face 收集模型卡片元数据以促进多学科人工智能研究：一个数据集。载于《第14届数据科学、技术与应用国际会议论文集》（2025），第583–590页。[35] Tan, X., Li, T., Chen, R., Liu, F. 和 Zhang, L. 使用预训练模型面临的挑战：从业者视角。载于 SANER'24（2024），IEEE，第67–78页。[36] Zhao, Z., Chen, Y., Bangash, A. A., Adams, B. 和 Hassan, A. E. 机器学习资产管理挑战的实证研究。《实证软件工程》29卷4期（2024），98。[37] Zhou, D.-W. 和 Ye, H.-J. 模型复用的统一视角：从小型到大型预训练模型。[38] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. 和 He, Q. 迁移学习综合综述。《IEEE 论文集》109卷1期（2020），43–76。