

国外人工智能开源生态运营模式剖析

何 婷, 徐 峰

(中国科学技术信息研究所, 北京 100038)

摘 要: 开源正成为新一代人工智能发展的全球趋势。本文从开源政策、开源组织、开源基础设施三个方面, 基于典型实践案例, 总结了国外人工智能开源生态的运营模式和主要特征。研究发现, 欧美等发达国家的人工智能开源生态已进入到一个元素基本完整、运营基本顺畅的发展阶段, 并通过所有权中立、社区治理专业化、顶尖企业引领、商业模式创新等方式实现开源生态的可持续运营。建议围绕打造良好的人工智能开源政策环境、推动开源组织有序分工合作、夯实开源基础设施等方面, 进一步完善我国人工智能开源生态。

关键词: 人工智能; 开源生态; 开源基金会; 领军企业

中图分类号: TP18; G321 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2022.01.010

开源 (Open Source) 是信息技术领域一种大规模协作开发的创新方式, 它改变了信息技术产业格局和商业模式。开源模式凭借强大的资源汇集和协同创新优势, 已成为推动人工智能技术快速迭代的关键动力和新一代人工智能发展的全球趋势^[1]。在国家战略层面, 人工智能的开源发展受到各国政府的高度关注。美国在《人工智能研发战略规划 2019 更新版》中强调要大力推进开源软件和工具; 英国《AI 路线图》提出鼓励开源和标准化的软件和工具; 意大利《国家人工智能战略》指出要构建一个开源的人工智能软件和数据库。在产业发展层面, 以谷歌、亚马逊、Meta、微软等为代表的顶尖互联网企业, 正凭借自身的数据、技术和资本等优势, 抢占人工智能开源平台的技术和规则主导权。

当前, 我国在全球人工智能开源领域中的参与度和贡献度不断提升, 但也面临着本土开源生态不完善、国际影响力不足、对外依赖度较高等问题。我国人工智能开源生态还处在早期成长阶段, 核心的开源框架主要依靠国外, 开源社区、

代码托管平台和开源基金会等基本要素, 发展尚不成熟, 不同开源主体之间的分工合作模式还在构建中。

针对人工智能开源这一话题, 国内外研究都较少。国内研究主要以高校、科研院所等发布的研究报告为主。上海交通大学编写的《激活 AI 创新之源——上海人工智能开源开放发展报告》重点介绍了目前国内外主要的人工智能开源开放平台及项目。国家人工智能标准化总体组的《人工智能开源与标准化研究报告》基于典型的人工智能开源项目分析了各开源技术的特点。中国人工智能开源软件发展联盟重点研究了人工智能开源软件的发展现状和应用案例。综上, 现有文献更侧重于从技术层面梳理国内人工智能开源的发展现状, 而从体制机制层面研究如何实现我国人工智能开源生态可持续运营的研究较为缺乏。鉴于此, 本文尝试从国外人工智能开源实践中总结出实现人工智能开源生态可持续发展的关键要素, 为更好地构建我国人工智能开源生态提供参考。

第一作者简介: 何婷 (1996—), 女, 研究实习员, 主要研究方向为科技政策、产业经济。

项目来源: 中国科学技术信息研究所创新研究基金青年项目“国外人工智能开源生态运营模式剖析”(QN-202109)。

收稿日期: 2021-11-08

1 人工智能开源生态的内涵

开源文化起源于国外的软件行业, 至今已有半个世纪的发展。开源一词来源于“开源软件”, 由开源软件促进会 (Open Source Initiative, OSI) 于 1988 年正式提出, 指的是开放源代码并遵循开源许可证, 可进行自由发布、自由复制、自由修改、自由使用的软件^[2]。随着开源软件的发展, 开源已成为一项全球性的流行运动^[3]。从本质上看, 开源是一种分布式协作开发模式, 一种大规模群体依托互联网平台, 通过共同参与和协作, 不断累积智慧、实现持续创新的方法。实践表明, 开源已经成为推动人工智能、区块链等新兴信息技术创新和产业发展的有效路径。其作用主要体现为: 集聚全球智力资源, 推动技术突破; 大幅减少重复性工作, 节约研发资源和时间成本, 加快技术迭代; 降低技术研发门槛, 加快新技术的推广和应用^[4]。

关于人工智能开源, 从开源内容看, 根据国家人工智能标准化总体组发布的《人工智能开源与标准化研究报告》, 人工智能开源全栈包括基础设施、引擎、应用使能以及人工智能业务四个层次, 既包括人工智能芯片、数据集、训练平台等基础设施以及知识图谱、深度学习和机器学习框架等核心技术, 还包括推荐系统、智能问答、语音识别等应用项目^[5]。

“开源生态”是“开源”和“生态系统”的组合词。“生态系统”最初是生物学领域的一个概念, 指的是生物体之间以及生物体与无机环境之间通过复杂交互作用所形成的统一整体^[6]。关于“开源生态”的概念, 目前业界和学术界并没有规范的定义。Jullien 等^[7]认为, 开源生态系统是开源项目、企业、开源社区以及独立开发设计者等不同角色之间的复杂网络关系。有研究者认为, 开源生态包括五个必备要素, 分别是开源项目贡献者和使用者 (人)、开源代码托管平台和开源组成分析工具等 (机)、源代码 (料)、开源许可证和社区管理办法等开源生态规则 (法)、开源社区 (环)^[8]。另有研究指出, 开源生态以开源项目为中心构建, 依托开源社区协作形成软件、硬件等开源项目; 涉及开源贡献者、开源使用者、开源运营者、开源服务者多重角色, 包含开源治理、

开源运营、开源商业等多个环节^[9]。

综合上述概念, 本研究认为, 开源生态是一个汇聚多方资源的技术创新生态圈。人工智能开源生态是以人工智能开源项目为中心, 由开源贡献者、开源服务者、开源使用者、开源运营者等参与主体组成的彼此依存、相互影响、共同发展的人工智能技术创新生态圈。其中, 开源贡献者主要指最初贡献开源项目的企业或个人, 开源使用者指开源项目的最终用户, 开源运营者主要指促进开源贡献者相互协作的主体, 开源服务者主要指负责为开源项目开发提供相关辅助服务以及为规范开源生态发展而制定相关政策或规则的主体。人工智能开源生态的运营模式就是上述开源参与者之间、开源参与者与开源项目之间产生联系的路径, 这种路径决定了人力、资金、基础设施等要素的投入方式、投入质量、投入规模和投入周期, 显著影响开源项目的成长和商业价值的传递。

2 国外人工智能开源生态的运营模式

通过对人工智能开源实践的梳理和分析, 可以发现, 由于开源技术发展较早, 整体开源环境较为成熟, 欧美等发达国家的人工智能开源生态已经进入到一个元素基本完整、运营基本顺畅的发展阶段 (见图 1)。不断完善的开源政策为企业投身人工智能开源创造了有利的外部环境, 并推动政府从开源项目的使用者向贡献者转变。开源代码托管平台、开源许可证等为人工智能开源提供了标准化的工具和规则, 保障开源的规模化、规范化发展。各开源主体间的良好互动支撑了人工智能开源生态的可持续发展。一方面, 企业、研究机构或个人作为开源项目的发起者或参与者, 自发组建起开源社区, 开源基金会作为开源项目的主要运营者, 负责为社区运营和项目管理提供专业化、多样化、全周期的服务; 另一方面, 企业与开源基金会之间又形成了互利共赢的合作机制, 企业以会员的身份向开源基金会捐赠项目和资金, 并影响或主导着基金会的发展, 而基金会则基于资源集聚的优势, 能帮助企业的开源项目发展成熟, 同时给予企业接触外部优质项目和挖掘优秀人才的渠道^[10]。

2.1 开源政策

欧美国家在部署人工智能发展时十分重视人

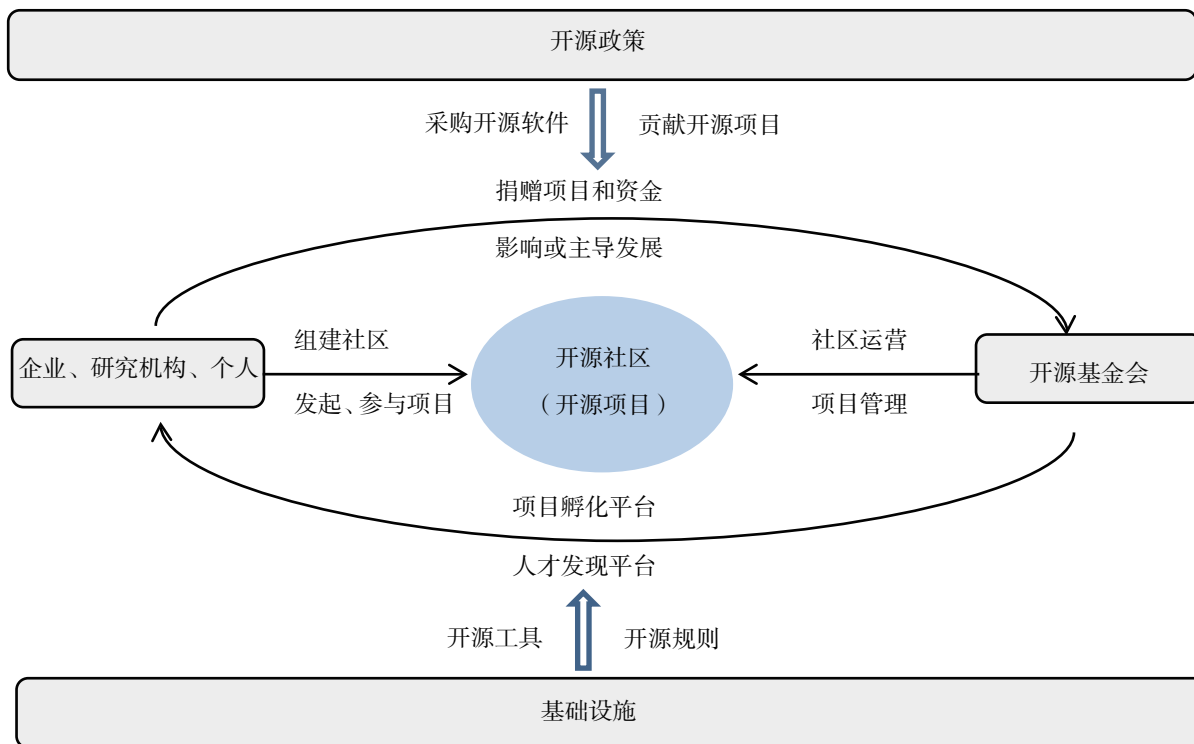


图 1 国外人工智能开源生态运营模式

工智能开源问题，强调通过政府采购、督促政府部门主动开源等措施推动人工智能开源。美国在2016和2019年的《人工智能研发战略计划》中均提出要开发开源软件库和工具包，要求政府部门不但要支持和使用人工智能开源技术，还要为开源项目贡献算法或软件；《2018美国国防部人工智能战略概要》指出开源社区是创新思想与人才的孵化器，国防部要加强与开源社区的合作，一方面要为开源社区提供数据、研究、技术以及挑战赛等支持，另一方面要积极参与开源生态的构建，发挥国防部在吸引人工智能人才、突破基于人工智能的国防技术等方面的作用。英国《AI路线图》（AI Roadmap）提出，应鼓励开源和标准化的软件和工具；《公共部门人工智能使用指南》（A Guide to Using AI in the Public Sector）要求确保政府部门使用的人工智能系统代码是开源的。

除了重视开源在人工智能技术发展中的重要作用之外,欧美国家还早在本世纪初就开始将推动开源软件发展作为一项独立的战略行动,通过支持研发、强制使用、优先采购以及推广应用等措施持续推动开源软件的发展^[11]。从2004年起,美国就

出台了一系列指导政府部门使用开源软件的政策。例如，2004 年美国行政管理和预算局（Office of Management and Budget, OMB）发布备忘录，要求增加对开源软件的采购。2016 年，美国行政管理和预算局正式发布《联邦源代码政策》，要求被联邦政府资助的网站、应用程序和其他软件项目发布源代码，并且各机构需要根据三年试点计划向公众发布至少 20% 的代码。

欧洲方面，2000 年欧盟委员会首次制定了开源软件战略，此后又分别于 2010 年、2014 年和 2017 年进行了三次更新，致力于推动开源软件在欧盟成员国政府内部的使用。英国 2010 年正式发布《政府的开源战略》（An Open Source Strategy for Government），提出在开源和非开源软件产品没有显著成本差异的情况下，应根据灵活性优先选择开源软件产品；在 2012 年发布的《采购政策注明 8/11：采购开源信息技术》、2013 年发布的《政府服务设计手册》等文件中，均强调政府采购应首选开源软件；在 2019 年发布的最新版《数字服务标准》中又要求政府开放新的源代码，并选择合适的许可证开源。法国在第二个国家开放科学计划中首

次将开源作为科学研究的重要组成部分, 提出支持公共资助的研究计划在开源许可下传播源代码, 重视高校和科研院所的源代码生产以及制定开源软件政策等措施。

2.2 开源组织

2.2.1 开源社区

开源社区 (Open Source Community) 是开源项目的主要开发组织, 一般指以开源项目为核心, 由地缘空间分散但拥有共同兴趣爱好的开发者根据相应的开源软件许可协议, 以民主、合作的形式, 共同进行项目开发、维护、增强等知识创造与传播的网络平台^[12]。开源社区的功能主要包括为开发者和用户提供交流平台、开源资讯、代码托管以及软件下载等。与传统专有软件开发组织 (主要是商业企业) 相比, 开源社区具有组织管理松散、没有明确的层级结构、交流民主等特点, 是一种开放式组织。

目前, 国外在计算机视觉、自然语言处理、深度学习等人工智能技术领域都存在专门的开源社区。这类社区通常专注于某一特定项目的共同研发。比如大家熟知的深度学习框架 TensorFlow, 它既是一个开源项目, 也是一个开源社区。为不断完善 TensorFlow, 谷歌搭建起 TensorFlow 社区, 吸引来自全球各地的开发者和使用者通过网络平台以及自发组织的线下活动, 进行交流互动、协作开发, 现已形成 70 多个用户组、170 多个机器学习谷歌开发者专家组和 12 个特殊兴趣组。OpenNLP 既是 Apache 软件基金会运营的一个自然语言处理开源项目, 也是一个开源社区, 其官方网站上不仅提供用于自然语言处理任务的模型、工具包等资源, 还设有项目任务开发专栏, 吸引更多开发者参与其中。由英特尔开发的计算机视觉项目 OpenCV 也基于网络平台开设了用户问答论坛以及贡献者通道, 推动 OpenCV 的应用和研发创新。

2.2.2 开源基金会

开源基金会是专门为支持开源项目而组建的非营利性组织, 旨在为开源项目提供资金、技术、市场等全方位支持, 为社区的建设和运营提供指导。成功的开源基金会一般都有一整套较为完整的服务体系, 以满足项目或社区在不同生命周期的需求。项

目初期, Linux 基金会、Apache 软件基金会 (ASF)、云原生计算基金会 (CNCF) 等主流基金会均有非常成熟的项目孵化流程, 给新项目提供技术研发、知识产权管理、社区建设等方面的指导。项目日常运营过程中, 基金会一方面会提供技术开发必要的基础设施支撑, 包括源代码管理、代码审查、问题追踪以及通信工具, 另一方面会借助大型会议、媒体、社交网络等途径宣传项目, 吸引更多开发者和用户参与项目维护。在资金来源方面, 开源基金会的日常运营主要依靠会员企业的会费, Apache 软件基金会、Linux 基金会、OpenStack 基金会 (现更名为开源基础设施基金会) 最高等级的会员每年会费分别是 12.5 万美元、50 万美元、35 万美元。

从国外实践看, 人工智能领域的开源基金会按业务范围可分为两类, 一类是设立人工智能子基金或为人工智能开源项目提供孵化服务的老牌综合类基金会, 如 Linux 基金会、Apache 软件基金会, 还有一类是专注于特定技术领域或项目的新兴基金会, 比如 Autoware 基金会。

Linux 基金会在 2018 年发起了一个专注于人工智能方向的子基金会 LF Deep Learning Foundation (2019 年更名为 LF AI Foundation), 旨在进一步推动人工智能和机器学习的发展。目前, 该基金会已拥有腾讯、AT&T、诺基亚、华为、中兴、百度、爱立信、IBM、滴滴、红帽等 50 家公司会员, 正孵化和已孵化出的开源项目共计 313 个, 覆盖机器学习、深度学习、强化学习、分布式计算、自然语言处理等多个领域。

Apache 软件基金会已为超过 350 个开源项目提供支持, 涵盖人工智能、大数据、云计算、物联网和边缘计算等众多技术领域。该基金会支持的人工智能典型开源项目主要有 Apache SINGA、MXNet、Mahout 等。其中, 由网易研究院和新加坡国立大学发布的开源分布式深度学习平台 Apache SINGA 是基金会资助的第一个深度学习项目, 可通过不同的运算符 (神经网络层) 构建深度学习模型, 已广泛应用于科研、医疗、金融等领域; MXNet 是由亚马逊联合华盛顿大学构建的深度学习框架, 2018 年进入 Apache 孵化器, 现已成为基金会的顶级项目; Mahout 是一个开源算法库, 集成了一些

可扩展的机器学习领域经典算法。

Autoware 基金会由开源协作组织 96Boards、日本智能驾驶技术公司 Tier IV、美国自主移动系统软件公司 Apex.AI 于 2018 年联合成立，致力于发展和资助自动驾驶开源协作项目 Autoware。该基金会是迄今为止全球自动驾驶领域唯一的开源组织，目前已有 22 家会员，其中包括华为和速腾聚创两家中国公司。

2.2.3 科技企业

开源的公益性质似乎和企业追求商业化的目的相悖，但面对开源软件不可阻挡的发展趋势，出于建立标准、降低成本、战略考虑和实现兼容性等动机，传统的 IT 企业和新兴的平台企业都纷纷推进自己的开源战略^[13]。GitHub 2018 年的年度报告显示，贡献者数量最多的十大机构中，排名第一的是曾公开反对开源软件的微软，其他还包括谷歌、红帽、英特尔、Meta 等科技企业以及加州大学伯克利分校、麻省理工学院、斯坦福大学等高校。虽然任何人都可以参与到开源项目开发中，但由于具有强大的号召力、稳定的研发团队以及持续的资金投入，相比独立的个体程序员而言，科技企业在人工智能开源生态中的贡献度和影响力更大。

在此轮人工智能浪潮兴起之际，开源就成为众多企业的首选。以谷歌、IBM、Meta、微软为代表的国际顶尖科技企业，凭借自身的数据、技术和资本等优势，持续在人工智能开源领域发力，主要通过三种方式参与人工智能开源生态的建设：一是将自行开发的人工智能软件实行开源。2015 年以来，美国人工智能研发巨头先后宣布开源自研的人工智能平台、引擎和工具包。Meta 开源了一组基于 Torch 的深度学习工具；谷歌将其开发的深度学习引擎 TensorFlow 开源，且据谷歌官方统计，截至目前，谷歌已开源了 50 个人工智能项目；微软开源了分布式机器学习工具包 DMTK；IBM 将机器学习平台 System ML 开源。二是通过开源基金会资助人工智能开源项目。IBM、谷歌、微软、Meta、亚马逊等企业均是 Apache、Linux、OpenStack 三家顶级开源基金会的高级会员，每年都会投入一定数量的资金支持开源基金会开展项

目孵化等服务。三是收购或投资开源平台或企业。2018 年，微软收购了全球最大的开源代码托管平台 GitHub，IBM 收购了全球知名开源解决方案提供商红帽，谷歌投资了全球第二大开源代码托管平台 GitLab。

2.3 基础设施

2.3.1 代码托管平台

开源既然要开放代码，让成千上万人协作，就需要一个足够友好和安全的代码存储环境。在这一背景下，代码托管平台应运而生。目前，全球使用率最高的代码托管平台主要有 GitHub、GitLab 等。经过十多年的发展，GitHub 已发展成全球最大的代码托管平台，汇集了 5 600 多万的开发者。随着人工智能的发展，GitHub 上以“深度学习”“自然语言处理”“机器学习”等为主题的存储库数量逐渐增加，存有大量开源框架、算法、工具包以及数据集，为人工智能解决方案的开发提供了基础支撑。据经济合作与发展组织（OECD）统计，2010—2017 年间，GitHub 上人工智能开源项目提交数占项目提交总数的比重从 0.26% 上升至 0.74%，增长幅度大约是其其他开源项目的 3 倍^[14]。

2.3.2 开源许可证

开源许可证是对开源核心理念的一种法律保障，保证开源项目能够合法地被自由使用和共享^[15]。开源许可证是一种授权许可，允许用户在承认原作者著作权的基础上拥有自由复制、修改以及再发布的权利。虽然各开源许可证对相关责任义务的约束度不同，但共同点主要包括：（1）承认著作权，要求署上原作者或版权持有人的姓名，这也是开源许可证最基本的限制；（2）允许私人使用和商业使用；（3）允许修改及修改后再发布；（4）免责声明，原作者或版权持有人不承担软件使用后的风险及产生的后果^[16]。可以说，开源许可证是解决知识产权专有性和源代码共享二者之间矛盾的关键。对于开发者而言，开源许可证赋予了其发表权和署名权等权利，起到一定的激励作用；对于使用者而言，开源许可证可以帮助其避免侵权等可能的法律纠纷；对于整个开源生态而言，开源许可证规定了生态参与者的权利和义务，是生态运行的核心规则，促进了开源项目的规范化发展。

以主流的深度学习框架为例(见表1), 目前人工智能开源项目主要使用 MIT 许可证、BSD 许可证、Apache 许可证等使用度最高也是最利于开源项目商业化的开源许可证^[17]。这些较为成熟的开源许可证为人工智能开源项目的开发和再传播提供了规范的法律环境和良好的商业环境, 而这背后则离不开开源基金会、科研院所等机构的智慧。MIT 许可证是麻省理工学院制定的, BSD 许可证是加州大学伯克利分校针对自研的 Unix 操作系统而推出的, Apache 许可证则是由 Apache 基金会在 MIT 许可证和 BSD 许可证基础上修改发布的。

表 1 主流深度学习框架使用的开源许可证

深度学习框架名称	所使用的开源许可证
MXNet	Apache 许可证
Caffe	BSD 许可证
TensorFlow	Apache 许可证
Torch	BSD 许可证
Theano	BSD 许可证
CNTK	MIT 许可证

资料来源: 根据公开信息整理。

3 国外人工智能开源生态的主要特征

如何组织并激励规模庞大但空间分散的开发群体进行持续创新是影响人工智能开源生态可持续发展的关键因素。在厘清运营模式的基础上, 笔者尝试从体制机制的角度梳理出国外人工智能开源生态的主要特征, 探究推动其逐步发展壮大关键性力量。

3.1 通过所有权中立促进各利益主体合作

“为他人做嫁衣”的忧虑是妨碍商业企业参与开源的原因之一。而通过开源基金会这一第三方机构, 实现所有权中立, 则使企业之间的合作成为可能^[18]。开源基金会通过与贡献者签订会员协议、转让协议、许可协议或根据开源许可证本身等各种方式, 获得了项目的所有权, 并负责管理知识产权。因此, 对任何项目参与者而言, 所有权都是中立的, 这便使得所有人员都能平等参与, 尤其是避免了商

业公司之间的竞争关系。另外, 让中立的第三方拥有所有权, 可以降低责任风险。基金会作为法人实体, 充当着社区成员的保护盾, 保护成员免于承担合同或与基金会本身有关的责任, 还可以保护未参加特定活动的成员免受其他成员行为的影响(例如, 将侵权软件引入基金会拥有的软件)。

3.2 依托专业化的服务机构培育开源社区

开源社区虽然以自组织为典型特征, 拒绝封闭式组织的集权模式, 但依然需要一定的治理机制来管理各项事务, 以维持组织的存续。为平衡社区的过度松散性和商业公司的过度中心化, 国外开源社区普遍从自发化、内部化进入到机构化发展阶段, 采用基金会这一组织方式来进行社区治理^[19]。基金会不直接参与项目开发等技术性事务, 而是承担方向把控、组织协调、人员培训以及基础设施等支持性工作。借助基金会提供的专业化服务, 开源社区既能保留决策自主并专注于技术创新, 又能实现规范化发展。

3.3 借助顶尖企业的号召力壮大开源力量

有研究表明, 成功的开源社区需要相关个体或组织进行主导和牵引^[20]。从国外实践看, 顶尖企业依靠自身的技术实力和行业影响力, 正成为推动人工智能开源队伍不断壮大的关键力量。一方面, 人工智能开源市场逐渐呈现出大型平台企业间相互竞争的格局。目前最受欢迎的人工智能开源项目主要来源于谷歌、Meta、亚马逊、微软等科技巨头(见表2); Linux 人工智能专项基金会已孵化完成的7个人工智能开源项目均来自于 Uber、IBM、微软、腾讯等大型科技企业。另一方面, 顶尖企业对开源基金会的影响力逐渐扩大。Linux 基金会的25位董事中绝大多数都来自于大型科技企业的管理层, 比如高通技术公司副总裁 David Marr、IBM 开放系统开发副总裁 Jessica Murillo 以及英特尔软件业务战略总经理 Melissa E. Evers-Hood 等。虽然也有不少人工智能初创企业投身开源, 但顶尖企业显然能投入更多的资金、人才和基础设施, 支撑大型开源项目的开发, 同时由于实力背书, 其发起的开源项目也更能吸引外部开发群体的参与。

3.4 创新商业模式实现开源项目市场价值

能够商业化是推动开源软件可持续发展的

表2 GitHub 上最受欢迎的人工智能开源项目

人工智能开源项目名称	最初开发者
TensorFlow	谷歌
Keras	谷歌
PyTorch	Meta
Scikit-learn	谷歌
Caffe	加州大学伯克利分校
MXNet	亚马逊
CNTK	微软
Theano	蒙特利尔大学
Caffe2	Meta

资料来源：根据公开信息整理。

重要因素之一。Apache 软件基金会创始人 Brain Beelendorf 指出，利他主义的开源和利己主义的商业模式结合在一起才能为开源做贡献^[21]。商业化带来的市场回报对潜在开源参与者形成了一种天然的外部激励^[22]。从国外实践看，当前人工智能开源项目的商业模式主要是通过建立生态圈拓展盈利空间，即发挥开源业务的“引流”作用，当开源项目在开发人员中变得流行甚至成为行业标准时，再逐步构建起人工智能产业链。比如，谷歌已将 TensorFlow 部署在几百万台设备上，占有深度学习框架市场最大份额，同时不断加强 TensorFlow 与自身云计算、芯片等业务的关联，比如推出了在谷歌云平台上运行 TensorFlow 的付费服务，开发出专门针对 TensorFlow 的人工智能芯片。

4 国外人工智能开源生态构建对我国的启示

在政府、高校、科研院所和企业等产学研力量的共同努力下，我国已建立起一些致力于推动人工智能开源的平台和组织，包括依托行业领军企业建立的自动驾驶、城市大脑、医疗影像等新一代人工智能开放创新平台，新一代人工智能产业技术创新战略联盟、中国人工智能开源软件发展联盟等产业联盟，以及启智社区（OpenI）、确实社区（Trustie）、木兰开源社区等开源社区。同时，我国本土的人工智能代码托管平台和开源许可证等基

础设施也基本形成，包括由鹏城实验室推出的汇智开源平台 iHub、由开源中国（目前国内最大的开源技术社区）推出的码云（Gitee）等代码托管平台，以及由中国开源云联盟制定的中国首个开源许可证“木兰宽松许可证”。此外，我国开源基金会也进入起步阶段，国内首个开源基金会——开放原子开源基金会于 2020 年 9 月正式成立。

我国人工智能开源生态的构成要素已基本形成，但相比国外，整体生态还处在早期成长阶段，开源社区、企业、开源基金会等主要开源主体之间协同创新的方式并不明晰。因此，有必要借鉴国外人工智能开源生态的运营模式，聚焦打造良好的人工智能开源政策环境、推动开源组织有序分工合作、夯实开源基础设施等方面，推动我国人工智能开源生态的可持续发展。

（1）加强政策的主动性和引导性。

我国人工智能开源顶层设计正在不断完善，越来越重视开源社区、代码托管平台等的建设以及开源知识产权保护。但相较欧美国家实行的“优先购买”“主动开源”等措施，我国开源政策的激励性、引导性还不够强。为带动各界对人工智能开源项目的研发热情，建议增强政府的主动性，一方面加大现有科技计划对人工智能开源项目的支持力度，并推动在政府资助下形成的人工智能研发成果的开源，扩大人工智能开源供给；另一方面研究在政务、医疗、安防、教育等社会民生应用场景中率先使用国产人工智能开源软件，从需求侧激励国内企业加大自主研发。

（2）推动产业联盟和开源基金会并行发展。

国外开源基金会在人工智能开源生态的可持续运营中发挥着重要的组织支撑作用。在国内，更多的是由产业联盟扮演这一角色。与开源基金会相比，产业联盟组织较为松散，更侧重于为创新主体之间的信息交流和项目合作提供平台。而开源基金会基于法人地位，能为开源社区提供法律保障；利用稳定的会费收入，能为开源项目提供更专业化、规范化的服务；通过与企业之间形成的互利共赢模式，能推动开源生态的良性循环。要在人工智能开源参与者之间建立起更为牢固的合作纽带，既需要开源基金会这类第三方服务机构提供有效的机制保

障和程序化的管理服务, 也需要产业联盟作为产学研沟通与合作的桥梁, 开展开源技术行业推广、开源许可证制定、开源软件标准研制、开源人才培养等基础性、公益性服务。因此, 建议我国在继续发挥产业联盟作用的同时, 强化基金会制度, 借鉴国际主流开源基金会的发展经验, 进一步发展我国本土的人工智能开源基金会。

(3) 发挥领军企业的引领作用。

领军企业出于构建自身商业生态的需求, 对人工智能底层技术进行开源的意愿较强。同时基于强大的研发能力、资金保障和行业影响力, 其也能够主导开源项目的产生、成长和商业化。目前, 百度、腾讯等国内大型科技企业以及商汤科技、旷视科技等人工智能垂直行业的领军企业都纷纷推出了自研的人工智能开源项目。在此基础上, 可以借鉴国外人工智能开源社区以特定项目为核心的发展模式, 支持领军企业围绕某一项目建立起相应的开源社区, 利用自身的行业号召力吸引更多开发者加入, 探索龙头企业牵引型开源社区的建设。

(4) 发展本土代码托管平台和开源许可证。

代码托管平台是开源项目协作开发的基础工作环境, 开源许可证是项目开源的核心规则, 但目前两者基本都被美国垄断。代码托管平台呈现 GitHub 一家独大的格局, 常用的开源许可证均来自美国。在中美科技竞争日益激烈的背景下, 若国内在人工智能开源领域继续过度依赖美国的基础设施, 可能也会遭遇“断供”或法律风险。事实上, GitHub 已明确表示其企业服务器及用户上传的信息都要接受美国法律监管, 包括美国出口管制法; 而部分开源许可证也规定其适用法律为美国法。为保障我国人工智能开源环境的安全性、自主性, 需要发展壮大本土的代码托管平台, 鼓励企业优先选择 Gitee 等国内代码托管平台, 并推动国家科技计划下产生的人工智能创新成果适时在本土代码托管平台上开源。同时发挥产业联盟的作用, 研制并推广符合中国开源发展特点的许可证。■

参考文献:

- [1] Cheng C, Li B, Li Z Y, et al. Automatic detection of public development projects in large open source ecosystems: an exploratory study on github[EB/OL]. [2021-05-10]. <http://www.cs.rug.nl/search/uploads/Publications/cheng2018adp.pdf>.
- [2] 陆首群. 开源、创新和新经济[M]. 北京: 北京交通大学出版社, 2017: 20.
- [3] 王伟, 周添一, 赵生宇, 等. 全球开源生态发展现状研究[J]. 信息通信技术与政策, 2020, 311(5): 42-48.
- [4] 上海交通大学, 机器之心. 激活 AI 创新之源——人工智能开源开放发展报告[R/OL]. (2020-07-11) [2021-05-12]. <https://ai.sjtu.edu.cn/info/news/159>.
- [5] 国家人工智能标准化总体组. 人工智能开源与标准化研究报告[R/OL]. (2019-04-28) [2021-05-12]. <http://www.cesi.ac.cn/201904/5037.html>.
- [6] Tansley A G. The use and abuse of vegetational concepts and terms[J]. Ecology, 1935, 16(3): 284-307.
- [7] Jullien N, Stol K J, Herbsleb J. A Preliminary Theory for Open Source Ecosystem Micro-economics[M]. Singapore: Springer, 2019: 49-68.
- [8] 何宝宏. 开源法则[M]. 北京: 人民邮电出版社, 2020: 89.
- [9] 中国信息通信研究院. 开源生态白皮书(2020年)[R/OL]. (2020-10-16) [2021-05-10]. <http://www.caict.ac.cn/kxyj/qwfb/bps/202010/P020201211560823978082.pdf>.
- [10] Feller J, Fitzgerald B, Hissam S A, et al. Nonprofit Foundations and Their Role in Community-Firm Software Collaboration[M]. Boston: MIT Press, 2007: 393-413.
- [11] Center for Strategic and International Studies. Government Open Source Policies[R/OL]. (2010-04-16)[2021-07-15]. https://csis-website-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/publication/100416_Open_Source_Policies.pdf.
- [12] 陈光沛, 魏江, 李拓宇. 开源社区: 研究脉络、知识框架和研究展望[J]. 外国经济与管理, 2021, 43(2): 84-102.
- [13] Wichmann T. Firms' Open Source Activities: Motivations and Policy Implications[R/OL]. [2021-07-05]. <https://www.math.unipd.it/~bellio/FLOSS%20Final%20Report%20-%20Part%20-%20Firms%20Open%20Source%20Activities%20-%20Motivations%20and%20Policy%20Implications.pdf>.

- [14] Baruffaldii S, Van Beuzekom B, Dernis H, et al. Identifying and Measuring Developments in Artificial Intelligence: Making the Impossible Possible[R/OL]. [2021-07-13]. https://www.oecd-ilibrary.org/science-and-technology/identifying-and-measuring-developments-in-artificial-intelligence_5f65ff7e-en. pdf.
- [15] 何东杰, 宋昊, 王琪, 等. 开源许可证及其检测工具研究 [J]. 计算机应用与软件, 2018, 35 (6) : 28-35, 53.
- [16] 何宝宏. 开源法则 [M]. 北京: 人民邮电出版社, 2020: 111-112.
- [17] 中国信息通信研究院. 开源软件知识产权风险防控研究报告 [R/OL]. (2019-07-16) [2021-05-12]. <http://www.caict.ac.cn/kxyj/qwfb/bps/201907/P020190710343670915282>.
- [18] Hunter P, Walli S. The Rise and Evolution of the Open Source Software Foundation[J]. Journal of Open Law, Technology & Society, 2013, 5(1): 31-42.
- [19] De Laat P. Governance of open source software: state of the art[J]. Manage Governance, 2007(11): 165-177.
- [20] 闫俊周. 开源式创新研究综述与展望 [J]. 技术经济与管理研究, 2016 (12) : 41-45.
- [21] 陆首群. 开源、创新和新经济 [M]. 北京: 交通大学出版社, 2017: 25.
- [22] 王宇. 开源创新: 激励多元化与公共产品的私人有效提供 [J]. 当代财经, 2013 (10) : 36-45.

The Operation Mode of Foreign AI Open Source Ecosystem

HE Ting, XU Feng

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: Open source is becoming the global development trend of the new generation of Artificial Intelligence (AI). This paper combs foreign typical open source cases in AI, including open source policies, open source organizations and open source infrastructure, and summarizes the operation mode and characteristics of foreign AI open source ecosystem. It is found that the AI open source ecosystem in Europe, America and other developed countries has entered a development stage with complete elements and smooth operation and realized the sustainable operation through some designs such as ownership neutrality, community governance specialization, leading of top enterprises and business model innovation. It is suggested to improve China's AI open source ecosystem by focusing on creating a good policy environment, promoting the orderly cooperation of open source organizations, and consolidating open source infrastructure.

Keywords: artificial intelligence; open source ecosystem; open source foundation; leading enterprises