中国开源年度报告预览





王伟

X-lab 开放实验室 华东师范大学



CONTENTS 目录





整体概况

年度亮点

内容框架

开放协作

中国开源年度报告发展历史









2020 中国开源年度报告

2020

中国开源年度报告

2021 中国开源年度报告 (2021.1.1-12.31)

日文版中国开源年度报告 日本語版中国オープンソース年度報告書

2022

2016

2020

2020 中国オープンソース年度報告

2019



2021

























GitHub 2019 数字年报

摘要 :在开源日益重要的今天,我们需要一份建立在全域大数据基础上的相对完整、可以反复进行推演的数据报告(报告、数据、算法均需开 本报告中使用 2019 年全年 GitHub 日志进行统计,总日志条数约 5.46 亿条。

关键词: 开源、行为数据、开发者行为、Github、数字报告

二、内容框架



目录 Contents

01 大事记篇

一. 开源技术大事记

1.1 人工智能与大模型 1.2 操作系统与编程语言

1.3 硬件技术与物联网

1.4 Data Infra

1.5 云计算与基础软件

二. 开源生态大事记

2.1 各大厂纷纷裁撤开源人员 2.2 著名开源大佬生计艰难

2.3 知名开源项目陆续停止开发

2.4 自由软件基金会的 40 年风雨历程

4.3.2 Linux 恶意软件增长率飚至 50%

4.3.3 npm 供应链面临的新威胁: 清单之惑

4.3.4 Electron 惊现十级漏洞!

4.3.5 解决 2,900 个漏洞,Google 奖励 1,200 万美元

4.3.6 GitHub 增加 SBOM 导出功能,使其更易于符合安全性需求

4.3.7 OpenAI、谷歌微软等设立一千万美元 AI 安全基金

五. 开源商业大事记

5.1 早期阶段融资事件

5.2 中后期阶段融资事件 5.3 收并购事件

六. 开源教育大事记

6.1 开源教育实践活动逐渐丰富,项目课题制+创新竞赛制成为主流 6.2 本土开源教育理论支撑开始丰富、开源硬件融入基础教育

6.3 开源教育论坛兴起, 开源+教育圆越扩越大

6.4 开源人才培养与认证逐渐成为标准体系

6.5 企业参与开源教育,新型产学研合作模式兴起

6.6 高校开源教育课程体系日益完善,参与开源积极性提高

2.5 开源社区老龄化现象

2.6 来自中国开源的一些好消息

2.7 各国政策对于开源生态的影响

三. 开源治理大事记

3.1 社区治理

3.1.1 Rust 社区相关争议

3.1.2 红帽相关争议

3.2 风险治理

3.2.1 伦理道德与社会风险

3.2.2 法律合规风险: 大模型版权归属问题

3.2.3 法律合规风险:人工智能相关法律诉讼

3.2.4 法律合规风险:人工智能许可证问题

3.2.5 法律合规风险:人工智能法案

四. 开源安全大事记

4.1 最新趋势和挑战

4.2 开源安全的法律责任

4.3 2023 年一些重要的开源安全大事件

4.3.1 Log4j 漏洞复活

○3 商业化篇

一, 开源生态助力 AI 快速发展

1.1 预训练大模型迅速发展,开源功不可没

1.1.1 預训练大模型发展迅猛

1.1.2 开源力量助推 AI

1.1.3 大模型的三个层次

1.2 开源是助推基座模型发展的第二动力

1.2.1 供给侧: 集中力量, 促进研发

1.2.2 需求例: 降低门槛, 抢占市场

1.2.3 生态侧: 汇聚多元, 长久增长

1.2.4 开源大模型的商业化实现路径

1.3 AI 开发者工具开源已成为行业阶段性共识

1.3.1 开发者工具在 AI 产业链中发挥着重要作用

1.3.2 开发者工具开源有重要意义

1.3.3 开发者工具开源需重视生态搭建

1.3.4 开源开发者工具商业化路径探索

1.3.5 开发者工具侧开源的成功案例

1.4 AI 应用层开源工具百花齐放

1.4.1 应用层开源工具百花齐放

1.4.2 应用层开源的驱动因素

1.4.3 大模型应用层开源的市场现状

1.5 大模型开源商业化面临的挑战

1.5.1 技术高速发展,开源项目需要持续迭代以保持竞争力

1.5.2 抄袭 / 借整范围难以界定

1.5.3 社区参与者难以对模型迭代提供直接贡献

1.5.4 开源技术发展快, 后期更新成本高

二. 开源安全挑战

2.1 开源软件网络安全

2.1.1 开源软件安全漏洞会被利用造成严重后果

2.1.2 开源软件网络安全问题相对普遍

2.1.3 开源软件湿润风险的应对策略

2.2 开源许可证的可控

2.2.1 开源许可证是一种针对开源资源使用者的约束,类别丰富

2.2.2 使用开源资源不遵守许可证会产生侵权风险

2.2.3 开源大模型许可证很大程度上区别于传统许可证

2.2.4 保障许可证可控的方式

2.3 开源 AI 安全

2.3.1 开源 AI 对数据安全提出新的要求

2.3.2 开源 AI 大模型的大量使用引发对于社会伦理的思考

三. 开源项目资本市场情况

3.1 全球市场状况

3.1.1 2023 全球 VC 投资规模减小,但 AIGC 是万众焦点

3.1.2 全球开源融资情况

3.2 中国市场状况

3.2.1 中国股权资本市场发展情况概括

3.2.2 国内开源生态稳步发展

3.2.3 国内开源公司融资保持热度

02数据篇

指标介绍

一. 总体宏观洞察

1.1 基础事件

1.1.1 GitHub 全域事件趋势

1.1.2 GitHub 和 Gitee 的总事件数趋势比较

1.1.3 GitLink 事件数分析

1.2 活跃仓库

1.2.1 GitHub 全域活跃仓库数目趋势

1.2.2 GitHub 和 Gitee 总体活跃仓库活跃度趋势与对比

1.2.3 GitHub 和 Gitee 总体活跃仓库 OpenRank 趋势与对比 1.3 活跃用户

1.3.1 GitHub 总体活跃用户数量趋势

1.3.2 活跃用户地理分布与排名

1.4 开源许可证

1.4.1 使用开源许可证的仓库数量 1.4.2 开源许可证种类变化趋势

1.5.1 2023 年开发者使用编程语言榜单

1.4.3 使用开源许可证仓库数量变化趋势 1.4.4使用木兰系列许可证仓库数量变化趋势 1.5 编程语言

1.5.2 2019-2023年开发者使用编程语言趋势

二. OpenRank 排行榜

2.1 全球开源仓库 OpenRank 排名 2.2 中国开源项目 OpenRank 排名 2.3 全球企业 OpenRank 排名 2.4 中国企业 OpenRank 排名 2.5 全球基金会 OpenRank 排名 2.6 国家和地区 OpenRank 排名

2.7 全球开发者 OpenRank 排名

2.8 中国开发者 OpenRank 排名

三.企业洞察

3.1 近 10 年全球企业 OpenRank 演变图 3.2 近 10 年中国企业 OpenRank 演变图 3.3 中国企业在 GitHub/Gitee 平台上的 OpenRank 占比

四. 基金会洞察

4.1 全球基金会 OpenRank 趋势分析 4.2 全球基金会项目 OpenRank 趋势分析

4.3 基金会旗下中国项目 OpenRank 趋势分析

4.4 开放原子基金会旗下项目 OpenRank 趋势分析

五. 技术领域调察

5.1 六大技术领域近五年整体发展趋势 5.2 各領域 OpenRank 与活跃度 Top 10 项目近五年变化趋势

5.2.1 云原生 5.2.2 人工智能 5.2.3 大数据 5.2.4 粉桐店

5.2.5 前端 5.2.6 操作系统

5.3 2023 年各領域 OpenRank TOP 10 榜单

5.3.1 云原生 5.3.2 人工智能 5.3.3 大数据

5.3.4 数据库

5.3.5 前端 5.3.6 操作系统

六. 开源项目洞察

6.1 项目类型 6.1.1 不同项目类型数量比例

6.1.2 不同项目类型 OpenRank 加总比例 6.1.3 不同项目类型近 5 年 OpenRank 变化趋势

6.2 项目 Topic 分析 6.2.1 热门 topic

6.2.2 热门 Topic 的仓库总 OpenRank 趋势

6.3 数据库领域项目分析 6.3.1 数据库领域各子领域 2023 年 OpenRank 和活跃度榜单

6.3.2 数据库领域各子领域下项目近五年变化趋势 6.3.3 数据库领域各子领域下项目的开源象限图

6.4 生成式 AI 領域项目分析

6.4.1 生成 AI 各子领域近 5 年增长趋势 6.4.2 生成式 AI 領域项目 OpenRank 和活跃度 Top 10 变化趋势 6.4.3 2023 年生成式 AI 領域项目 OpenRank 和活跃度 Top 10 榜单

七. 开发者洞察

7.1 开发者的地区分布 7.2 开发者工作时间分析

7.2.1 全域开发者工作时间分布

7.2.2 项目工作时间分布

7.3 开发者角色分析 7.3.1 各角色数量分布

7.3.2 2023 年各角色新增情况

7.3.3 开发者演化视角 7.4 机器人账号分析

7.4.1 机器人活跃数据分析 7.4.2 机器人事件类型分析 7.4.3 机器人账号工作时间分布

八. 案例分析

8.1 openEuler 社区案例分析 8.2 中国开发者贡献的顶级仓库列表

7.4.4 GitHub 协作机器人事件数量排行榜

一. 报告背景

3.1. 受访者特征

3.1.2 所处行业、职业 3.2 开源参与情况

3.2.2 开源产品使用情况

3.2.4 信息交流

3.3.2 开源贡献方式 3.3.3 开源贡献内容 3.3.4 激励机制

3.4 社区运营调查

3.4.1 所在开源社区概况 3.4.2 开源社区管理

3.4.3 开源软件商业化调研 3.5 开源发展调研

3.5.1 开源发展态势

04 问卷篇

二. 问卷结果预览

三. 问卷分析

3.1.1 年龄、性别、学历、城市

3.2.1 开源社区的参与程度

3.2.3 技术方向

3.3 开源贡献情况 3.3.1 开源贡献参与程度 3.5.2 人工智能对开发者及开源生态的影响与挑战

问卷篇

- 调查对象:覆盖开发者、社区成员、贡献者、学生、政府企业管理人员
- 调查内容: 主要涵盖个人信息、工作状况、开源社区以及开发者技术等
- · 调查方法: 以在线问卷方式搜集样本和数据,交叉对比法分析数据
- **分发渠道**: 开源社、Apache Asia、第九届中国开源年会、 2024 开放原子开发者大会、2024 开源产业生态大会
- 问题类型: 单选、多选、开放性
- 问题数量: 总共四十题 (根据受访者的身份每人20~30题)
- 样本量: 预计 500+ 份



受访者特征

受访者年龄分布均匀,受教育程度普遍在本科以上,性别、 地区分布符合中国开发者地理分布;涵盖了多种计算机行业 职业身份。

开源参与情况

开源社区的活跃情况是受访者尤其关注的方面;**人工智能**成为了大多数受访者关注的技术领域。

开源贡献情况

受访的开源社区贡献者更多在**技术基础类型**的仓库贡献;受 访者贡献开源社区的动机多为**社区/荣誉激励**,对物质激励 的要求较低。

社区运营调查

大多数受访运营者所在开源社区**有专人负责社区运营**。近半 数受访者所在企业注重**开源软件的使用规范和管理**。

国内开源发展调查

受访者对国内开源未来发展持**乐观态度**。关于人工智能在开源生态的发展态势,开发者普遍看好其在**提高效率、自动化测试和数据分析方面**的应用前景,认为**数据安全、透明度、伦理问题**等是主要面临挑战。

数据篇: OpenDigger





- <u>OpenDigger</u> 是一个面向<u>开源数字生态</u>的一站式数据挖掘与信息服务项目,目标是构建开源领域的数据基础设施,促进开源生态的持续发展,**实现开源治理系列标准中的关键指标**。所支持的开源领域数据包括 GitHub / Gitee / AtomGit 行为日志数据、制品库数据、安全漏洞数据、社区问答数据等。
- · 代表性下游项目: <u>HyperCRX</u>、 <u>OpenLeaderboard</u>、 <u>OpenGalaxy</u>、 <u>OSGraph</u>、<u>OpenPerf、OpenTalent</u>等。

- ・GitHub 行为日志数 80 亿条
- ・Gitee 行为日志数 2600 万条 (部分)
- ・NPM/PyPI等制品库数据 620 万条
- ・CVE 安全漏洞数据 16 万条
- ・StackOverflow 问答帖 2500 万条
- · 标签类数据, 其中 GitHub 组织数量 413 个, 涉及仓库数 89427 个

指标

For repos

- [标准院/X-lab] activity
- ・ [标准院/X-lab] <u>openrank</u> 💮
- [标准院/X-lab] attention
- [标准院/X-lab] stars
- [标准院/X-lab] issue_comments
- [标准院/X-lab] participants
- [LF/CHAOSS] technical_fork
- [LF/CHAOSS] issues_new
- · [LF/CHAOSS] issues closed
- · [LF/CHAOSS] code change lines add
- [LF/CHAOSS] code_change_lines_rem
- [LF/CHAOSS] code_change_lines_sum
- · [LF/CHAOSS] change requests
- [LF/CHAOSS] change_requests_accept
- [LF/CHAOSS] change_requests_reviews
- [LF/CHAOSS] bus factor

For users

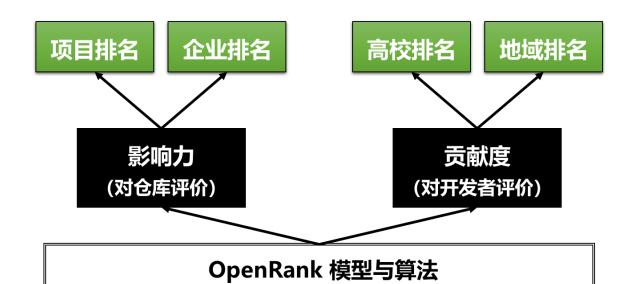
- [标准院/X-lab] activity
- [标准院/X-lab] openrank



覆盖开源软件	生态项目	实体数量	开源情报
2亿+	12 个	2.5 亿+	10 万+
仓库数量	开源评价场景	合作伙伴	参与社区
1500 万+	16 个	50 个	120 个
关系数量	企业采纳	榜单数量	CVE 漏洞数据
44 (Z+	500 家+	20 个	16 万+
开发实体数	知识产权	仓库容量	社交平台数据
1200 万+	110 个	1 PB	2500 万+
组织数量	开放 API	开源许可证	标签数据
413 个	120 个	3000+	10 万+

数据篇: OpenRank 开源评价标准体系



















中兴通讯、开放原子、浪潮信息、棱镜七彩、九州未来、中国软件评测中心、上海计算机软件技术开发中心、阿里云、蚂蚁集团、百度网讯、东软集团、中移软件、普元信息、华胜天成、东软集团、浪潮云、中移系统、金蝶天燕、谐云科技、中电标协等













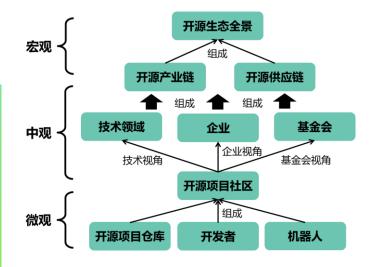


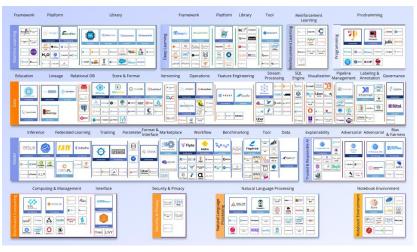


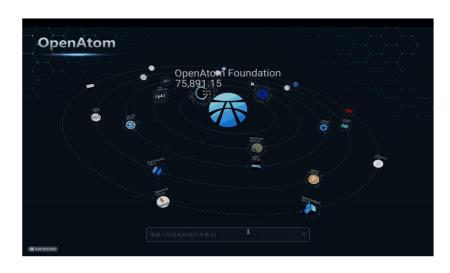
- ・ 发表学术论文超过 80 篇, 其中 CCF-A/B 类论文超过 20 篇
- ・专利 15 项
- ・软著 24 项
- · 开源评价学工作组获国际测试 委员会(BenchCouncil)立项

三、年度亮点



























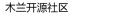


























年度亮点 1: 更全面的理论框架



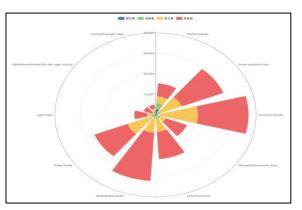
- · 总体宏观篇
- ・ 排行榜篇
- ・企业洞察篇
- · 基金会洞察篇
- · 技术领域篇
- ・ 开源项目篇
- ・开发者篇
- ・ 案例分析篇

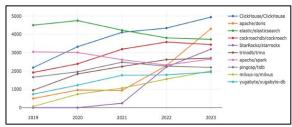
拟新增

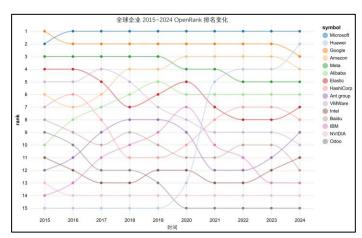
- ・ AIGC与大模型篇 (蚂蚁)
- ・ 商业洞察篇 (COSS)
- ・ 高校洞察篇 (OSPP)
- · 开源大赛篇 (开放原子)
- ・ 开源知识图谱篇 (OSGraph)
- · 开源时空洞察篇 (OSSInsight)
- ・ 开源战略事件篇 (CCF 开委会)
- AtomGit、GitCode、GitLink、 HuggingFace 等平台数据

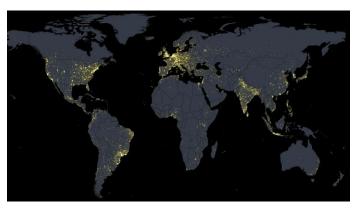


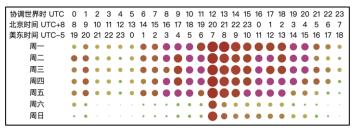
Ø	🖎 全球开源仓库 OpenRank 排行榜 (Top 20)										
	ŧ.	¥名称	OpenRank	参与人数	许可证	托管平台	国	R			
01		NixOS/nixpkgs	22133.13 🚧 4139.52	6,915	MIT License	Gattub					
62		home-assistant/core									
63											
64	•				Creative Commons Attribution 4.0 International						
05	0						•				
06		godotengineigodot									
07	ø,	elidianaandrade/dio-lab-open-source									
08	0						۰	Belgium			
		futter:flutter			BSD 3-Clause "New" or "Revised" License			United States of America			
10		microsoft/winget-pkgs					•	United States of America			
11		zephyrproject-rtos/zephyr									
12	•	elastic/kbana					•				
13							•	United States of America			
14	÷	mindspore/mindspore									
15	(3)										
16	1										
17	89										
18		grafana/grafana			GNU Affero General Public License v3.0	GitHub					
19	20		6968.53 🥕 3699.19			GitHub					
20											







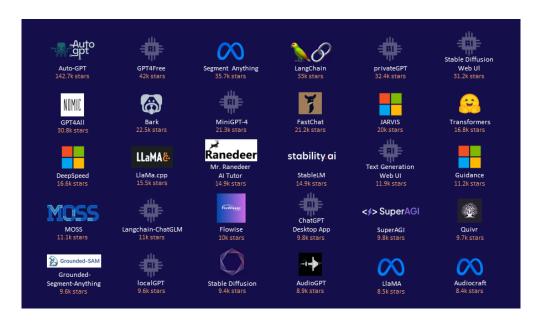


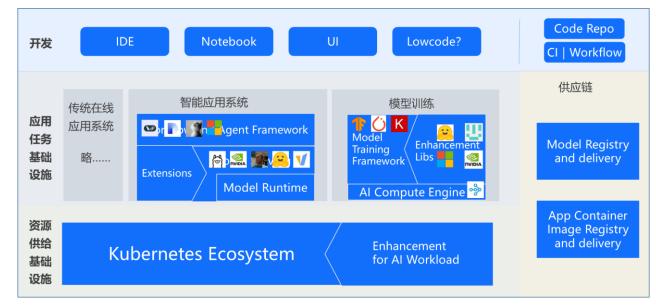


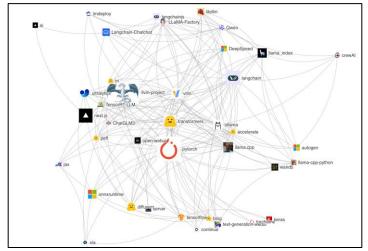
年度亮点 2: 更加深入的领域分析(人工智能) 🍎

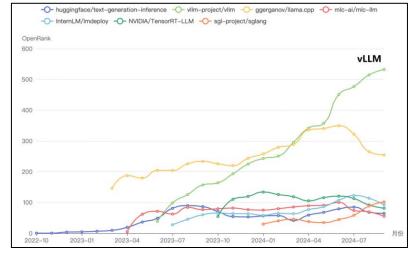


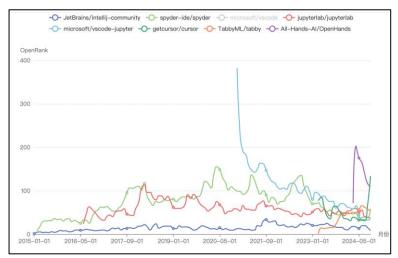












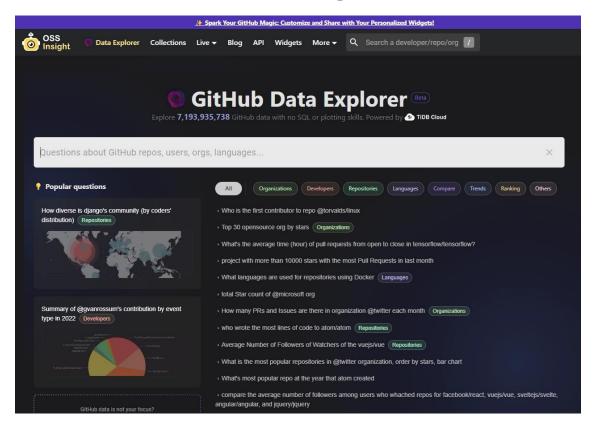
年度亮点 3: 更多样的分析方法 loss sight © OSGraph



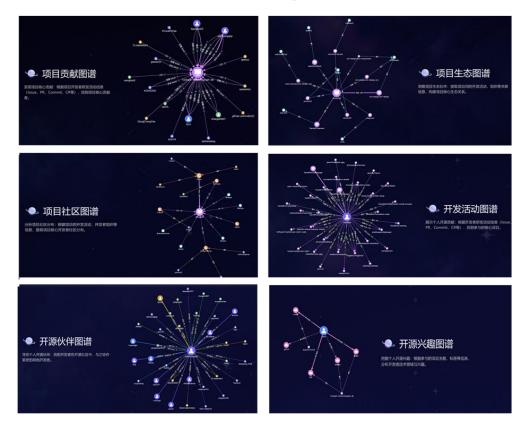




OSS Insight



OSGraph



年度亮点 4: 中国高校开源贡献全景图





- ・涉及全球 224 个社区
- ・2300多个开源项目
- ・分布全球高校 700 余所
- ・参与学生 6500 余人
- ·已解决问题/任务 1030





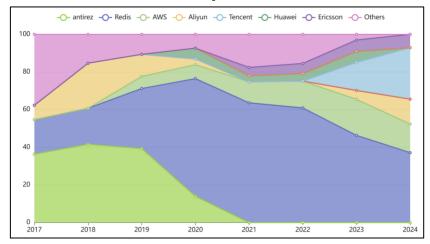
 43400位
 241场
 4980个
 573顷

 报名人数
 各城市举办百余场宣讲培训
 围绕各领域征集上干优秀作品
 产生百余项技术效果

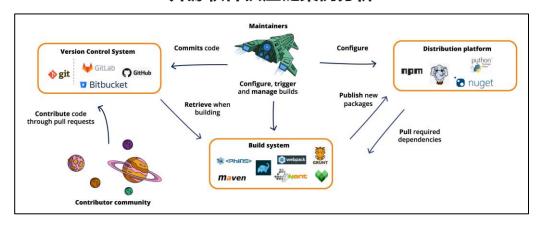
年度亮点 5: 更有趣的分析案例



Redis 开源社区 OpenRank 时序变化



开源软件供应链案例分析



openEuler 社区案例分析



开源学习社区案例分析





四、开放协作:连接与贡献





















































2024 中国开源参与情况问 卷调查



扫一扫或长按识别二维码

