

# 开源日报

OPENSOURCE DAILY



Issue: 26.01.2024

Date: 22.02.2024



总第014期

开源中国社区 - 数据分析



# ABOUT US

## 欢迎来到开源日报

您探索和理解开源世界的日常指南。在这里，我们每天从宏观到微观，全面展现开源技术和文化的动态图景。开源日报致力于捕捉这个不断演进的领域的脉搏，为您提供开源技术如何塑造我们的世界和未来的深度洞察。

无论您是资深技术专家、行业观察者，还是对开源文化充满好奇的新手，开源日报都将为您提供独到的视角和深度的洞察。我们相信，每一行开源代码背后都有一个故事，每一个项目都蕴藏着无限可能。

**OSChina 数据分析部**

加入我们的旅程，让我们共同揭开开源世界的奥秘，见证技术的力量如何每天以崭新的面貌展现在我们面前。

# 目录

<b>开源数据   OPEN SOURCE DATA</b> .....	<b>1</b>
开源社会数据 .....	1
开源环境指数 .....	2
开源话题情绪走势 .....	2
开源话题热词走势 .....	3
渠道传播TOP .....	3
今日活跃度渠道评估 .....	4
开源话题区域提及走势 .....	4
<b>今日要点   TODAY FOCUS</b> .....	<b>5</b>
StarkNet 撒钱，程序员在 Web3 领了 14 万 .....	5
谷歌发布轻量级开源大语言模型 Gemma.....	5
夜莺监控 V7 第一个 beta 版本来了 .....	6
<b>今日观察</b> .....	<b>7</b>
社交观察 .....	7
媒体观察 .....	8
<b>事件点评</b> .....	<b>9</b>
数百位名人签署公开信，呼吁制定反深度伪造立法 .....	9
StarkNet 撒钱，程序员在 Web3 领了 14 万 .....	10
<b>今日热门项目</b> .....	<b>11</b>
<a href="#">paperless-ngx/paperless-ngx</a> .....	11
<b>每日GitHub精选</b> .....	<b>12</b>
<b>每日Gitee精选</b> .....	<b>13</b>
<b>开源之声</b> .....	<b>14</b>
用户观点 .....	14
媒体观点 .....	15
<b>附：中国基础软件开源产业主要参与者</b> .....	<b>16</b>
<b>报告声明</b> .....	<b>19</b>

# 开源数据 | OPEN SOURCE DATA

## 开源社会数据

2024年02月21日 18:00:00~2024年02月22日 18:00:00, 共监测到信息总量9211条, 其中APP2302条 (占比24.99%), 网页2190条 (占比23.78%), 微信1328条 (占比14.41%), 微博1097条 (占比11.91%), 视频782条 (占比8.49%), 问答538条 (占比5.84%), 论坛508条 (占比5.52%), 头条号436条 (占比4.73%), 报刊30条 (占比0.33%)。



### 新闻平台

数据: 2190条 占比: 23.78%



### 社交平台

数据: 2425条 占比: 26.32%



### 视频平台

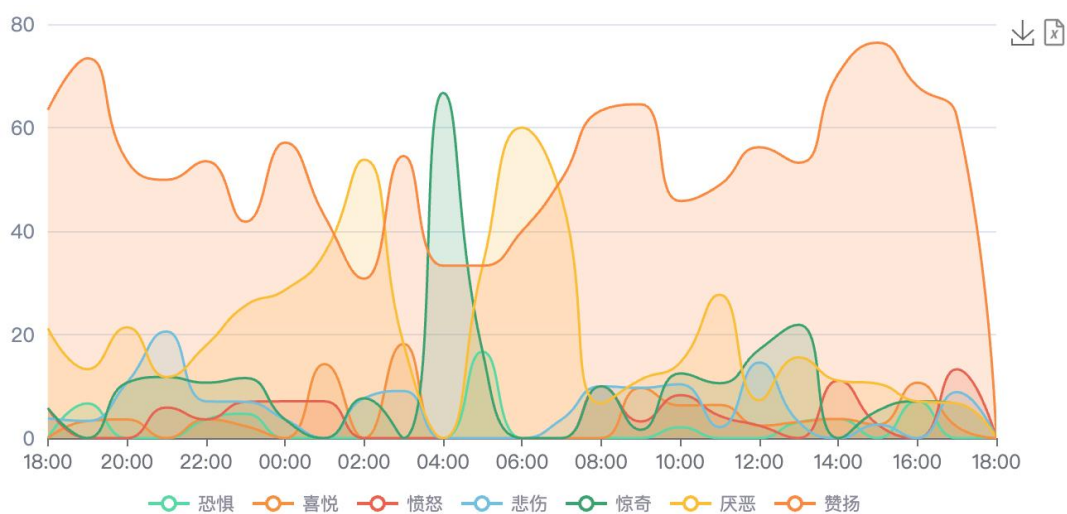
数据: 782条 占比: 8.49%



### APP平台

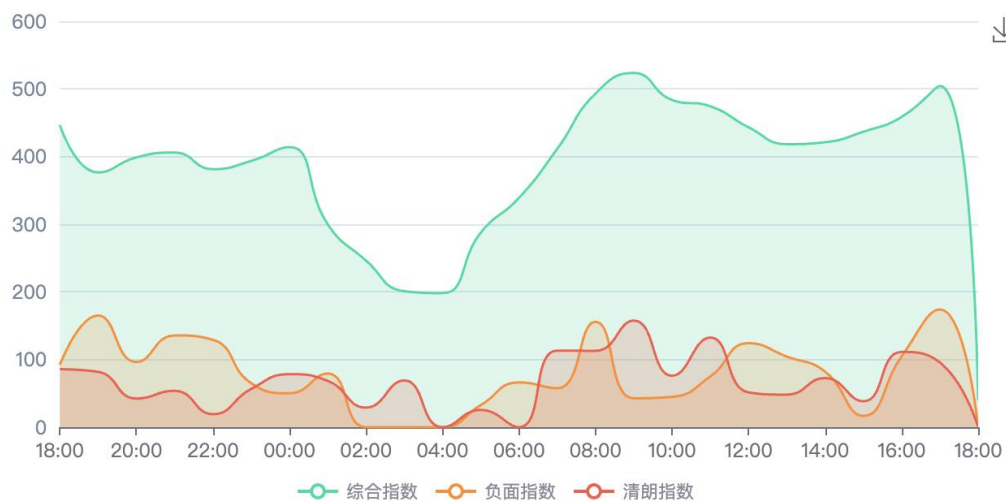
数据: 2302条 占比: 24.99%

## 开源话题情绪走势



今日开源环境舆论中，喜悦和赞扬占比6成左右，情绪一般

## 开源环境指数



舆情指数模型POI，根据舆情系统目前可采集数据设定。针对正面舆情跨平台传播热度进行评估（网页、微信、微博、客户端文章数量和正面舆情数量占比），暂时未考虑媒体影响力、事件敏感程度、历史风险性等其他要素。



地区	数量
北京	1300
上海	536
广东	430
山东	58
四川	58
浙江	56
安徽	54
河南	50
福建	40
江苏	36

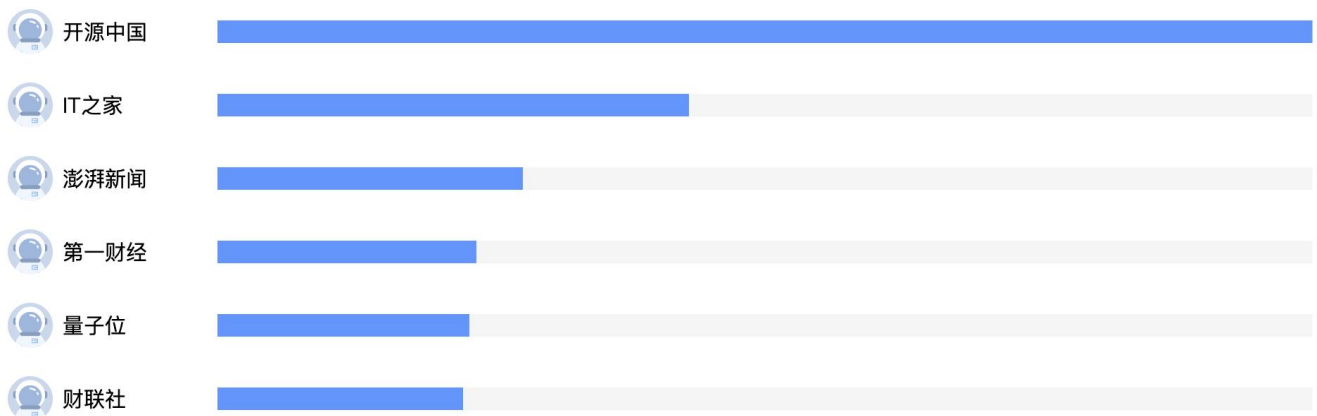


## 开源话题区域提及走势

分析时段内，发布信息地域信息地域主要集中在北京1300条（占比45.47%）、上海536条（占比18.75%）、广东430条（占比15.04%）、山东58条（占比2.03%）、四川58条（占比2.03%）。

今日北京、广东、上海地区对开源话题的兴趣度占比在 78% 左右。

## 今日活跃度渠道评估



# 今日要点 | TODAY FOCUS

## 谷歌发布轻量级开源大语言模型 Gemma

Gemma 是一款轻量级、先进的开源模型，供开发者和研究人员用于 AI 构建。Gemma 模型家族包括 2B（20 亿参数）和 7B（70 亿参数）两种尺寸，能够在不同的设备类型上运行，包括笔记本电脑、桌面电脑、IoT 设备、移动设备和云端。

### 编辑简评

谷歌发布的轻量级开源大语言模型 Gemma，标志着在AI领域持续发展的进程中，对于开发者和研究人员而言，提供了更灵活、更高效的构建AI应用的选择，对于推动人工智能技术的普及和应用具有重要意义。

## StarkNet 撒钱，程序员在 Web3 领了 14 万

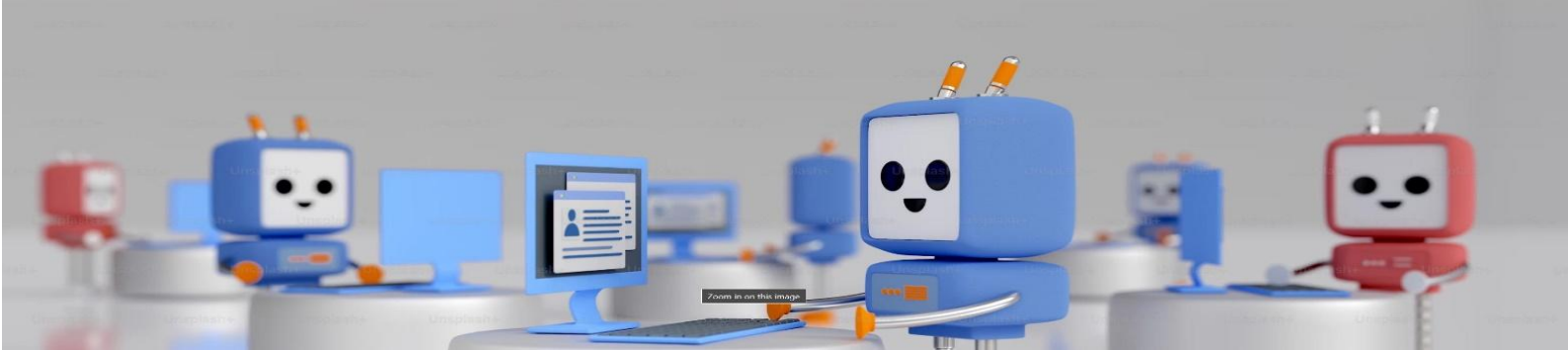
StarkNet 启动空投活动，GitHub 排名前 5000 开源项目的贡献者可领取价值 \$200 奖励。

- StarkNet 公链项目为了激励开发者参与其平台建设，启动了空投活动。
- 如果曾向 GitHub 上获得较多 Star 的项目提交过 PR，就有资格领取 111.1 STRK 的空投奖励。
- 只需要使用 OAuth 2.0 登录，就可以直接领取。

### 编辑简评

StarkNet启动的空投活动吸引了广泛关注，对于开源社区的贡献者来说是一次不错的奖励机会，有助于推动项目的发展和创新。这种激励机制对于推动开源社区的发展和项目的吸引力有着积极的作用。

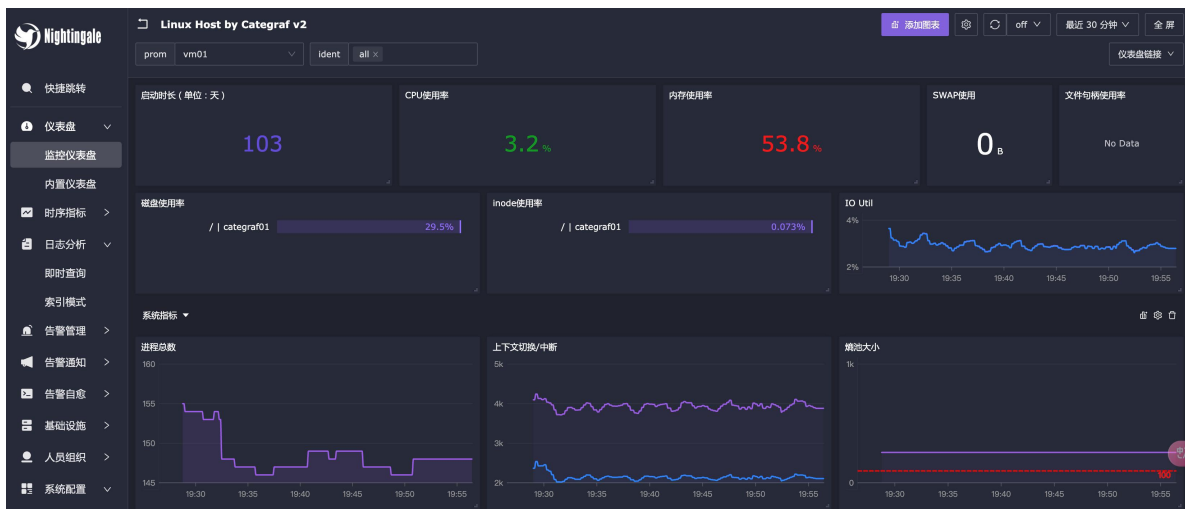




## 夜莺监控 V7 第一个 beta 版本来了

夜莺项目从 2024 年开始开发 V7 版本，重点做体验优化，V7 和 V6 版本兼容可以平滑升级（V6 升级到 V7 只需要替换一下二进制重启即可，如果是容器部署，只需要更新镜像并重启）。

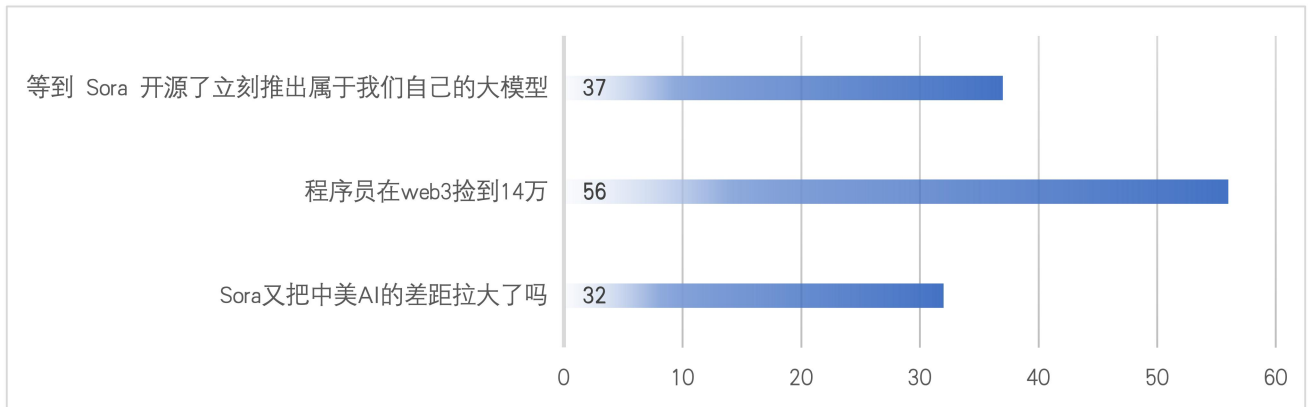
夜莺监控是一款开源云原生观测分析工具，采用 All-in-One 的设计理念，集数据采集、可视化、监报告警、数据分析于一体，与云原生生态紧密集成，提供开箱即用的企业级监控分析和告警能力。夜莺于 2020 年 3 月 20 日，在 github 上发布 v1 版本，已累计迭代 100 多个版本。夜莺的核心研发团队，也是 Open-Falcon 项目原核心研发人员，从 2014 年（Open-Falcon 是 2014 年开源）算起来，也有 10 年了，只为把监控这个事情做好。



### 编辑简评

夜莺项目的V7版本开发重点在体验优化，兼容性良好且易于升级，进一步巩固了其在开源云原生监控领域的领先地位，为用户提供更为便捷、高效的监控分析和告警能力。此外，V7版本开发意味着在云原生监控领域将有新的突破，重点优化用户体验，而与其与V6版本的平滑升级以及对云原生生态的紧密集成，将为用户带来更为便捷和高效的监控解决方案。

社交观察



关于Google开源Gemma的一些零散的看法

开源模型应该把主要精力放在开发并开源出性能足够强的小规模大模型上（SLLM，Small Large Language Model），我觉得谷歌的开源策略是非常合理的。目前看，作出强大的SLLM并没有太多技巧，主要是把模型压小的基础上，大量增加训练数据的规模，数据质量方面则是增加数学、代码等数据来提升模型的推理能力

- 微博 张俊林say

目前的人工智能技术远未达到人类智能水平，甚至连猫的智能水平都没达到

被视为深度学习领域“教父”之一的Meta（META）首席人工智能科学家Yann LeCun谈到了公众对强大人工智能模型的普遍恐惧，他认为这种恐惧大多是想象中的，因为目前的人工智能技术远未达到人类智能水平，甚至连猫的智能水平都没达到。尽管他确信人工智能最终会达到所谓的通用人工智能（AGI）阶段，但这一时间线可能比大多数研究者认为的要长得多。

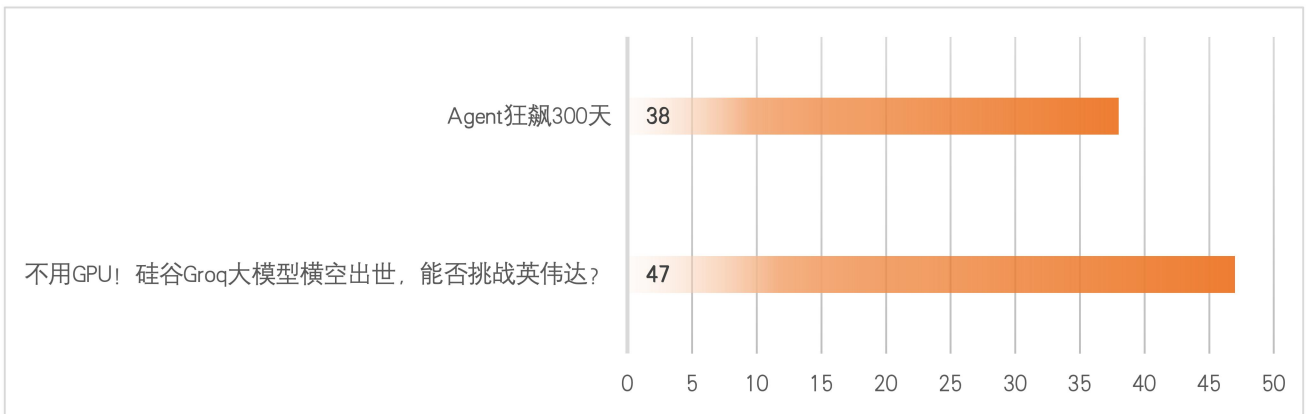
- 微博 高飞

李开复回应 AI 大模型使用美国开源技术：行业标准做法

就像“大多数其他AI公司一样”，零一万物的AI模型建立在LLaMA基础之上，使用开源技术是行业的一种标准做法。他指出，零一万物使用自己的数据和算法，从头开始训练其AI模型，这些才是其大模型“优越性能”的“主要决定因素”。

- 微博 凤凰网科技

媒体观察



### 为什么说Sora冲击的不只是抖音，还有一帮机器人公司？

Sora会创造比GPT更大的价值吗？它的问世对抖音、YouTube、派拉蒙等视频制作公司或分享平台意味着什么？为什么OpenAI和Meta在发布视频模型时都强调拥有「世界模型」的基础？什么是「世界模型」？「世界模型」对人工智能意味着什么？以及为什么是OpenAI做出了Sora而不是其他公司？

- 第一财经

### 谷歌发开源模型，不Open的只剩OpenAI

全世界都在为Sora惊艳，而谷歌却仍在默默地发语言模型。当地时间2月21日，谷歌推出了基于Gemini研究和开发的新型开源模型系列“Gemma”，并明确允许商业使用。谷歌发布开源Gemma之后，OpenAI成了唯一一个在这波AI热潮中没有发布过开源模型的AI公司。在Google Deepmind 联合创始人CEO Demis Hassabis的发布贴下，就有人@Sam Altman质疑OpenAI何时才能Open。

- 数智前线

### Google加入开源战局——Gemma来了！相当于OpenAI把GPT-3开源了

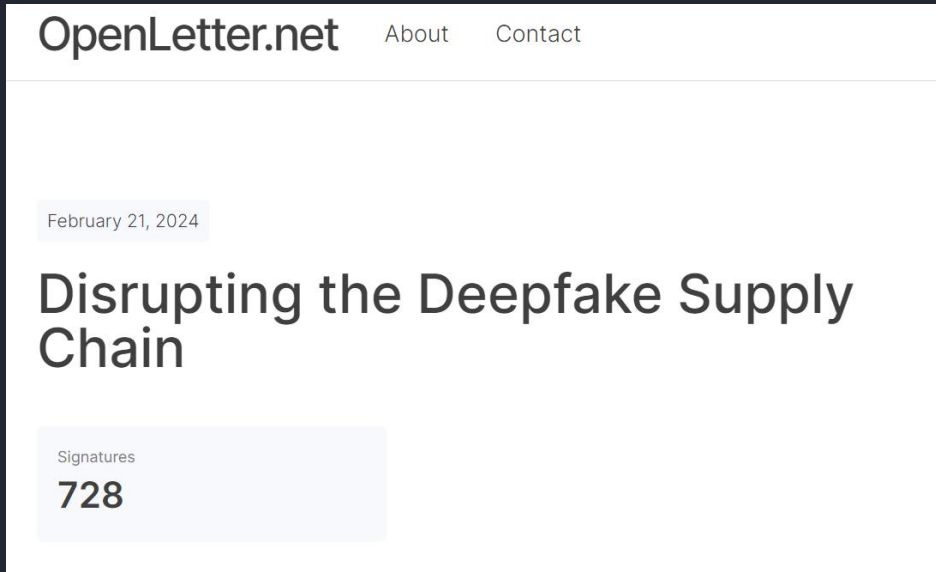
2月21日，在与闭源的OpenAI打得火热的同时，Google突然加入了开源的战局。北京时间夜间，Google突然宣布，开源了一个新的模型系列Gemma，这个模型使用了与它最强的Gemini同源的技术，并且在一系列的标准测试上秒杀了几款今天最热门的开源模型。怎么理解这个动作的重要性呢？你可以粗暴地理解为：

这有点像现在正在训练更强大的GPT-5的OpenAI，把GPT-3的低参数版给开源了。（前几天Sam Altman被问过这个问题，你们是否会在未来把GPT3开源，他没有直接回答。现在看来Google针对性很强啊。）？

- 硅星GenAI

# 事件点评

## 数百位名人签署公开信，呼吁制定反深度伪造立法



数百名 AI 界人士签署了一封公开信，呼吁严格监管 AI 生成的冒名顶替或深度伪造（Deepfake）内容。公开信指出，“深度伪造”是指未经同意或严重误导的人工智能生成的声音、图像或视频，合理的人会误以为是真实的。这不包括对图像或声音的轻微改动，也不包括容易识别为合成的无害娱乐或讽刺。如今，深度伪造通常涉及性图像、欺诈或政治虚假信息。由于人工智能发展迅速，使得深度伪造变得更加容易，因此我们需要为数字基础设施的运行和完整性提供保障。“Deepfakes 对社会的威胁日益严重，政府必须在整个供应链中施加义务，以阻止 Deepfakes 的扩散。”



事件点评 Event Search

这封公开信的发布对于引起社会对深度伪造问题的关注具有积极意义。这些专家的声音将有助于推动政府和产业界采取更严格的措施来应对深度伪造的威胁。

自身影响力

辐射影响力



+ Add funds

Send



**STRK airdrop is now LIVE!**

Click here to claim your STRK



**Ethereum**

0.0 ETH

**\$0.00**



**Starknet**

111.1 STRK

**\$199.98**

## StarkNet 撒钱，程序员在 Web3 领了 14 万

StarkNet 公链项目为了激励开发者参与其平台建设，启动了空投活动。如果曾向 GitHub 上获得较多 Star 的项目提交过 PR，就有资格领取 111.1 STRK 的空投奖励。只需要使用 OAuth 2.0 登录，就可以直接领取。

有程序员晒出自己在本次空投活动获得的奖励——近 14 万人民币。领取规则：

- 1、截止到 2023 年 11 月 15 日，至少对全球排名前 5000 的仓库提交过三次代码贡献。
- 2、其中至少有一次贡献是在 2018 年或之后完成的。



事件点评 Event Search

这次空投活动为开发者提供了一种额外的激励机制，鼓励他们积极参与到 StarkNet 的生态建设中。不仅有助于项目的发展，还有助于整个 Web3 生态的繁荣。但也有一些潜在的问题需要关注。首先，这样的空投活动是否会导致过度关注和短期炒作，而忽视了项目的长期发展？其次，如何确保空投奖励的公平性和透明度，避免不法分子的滥用？

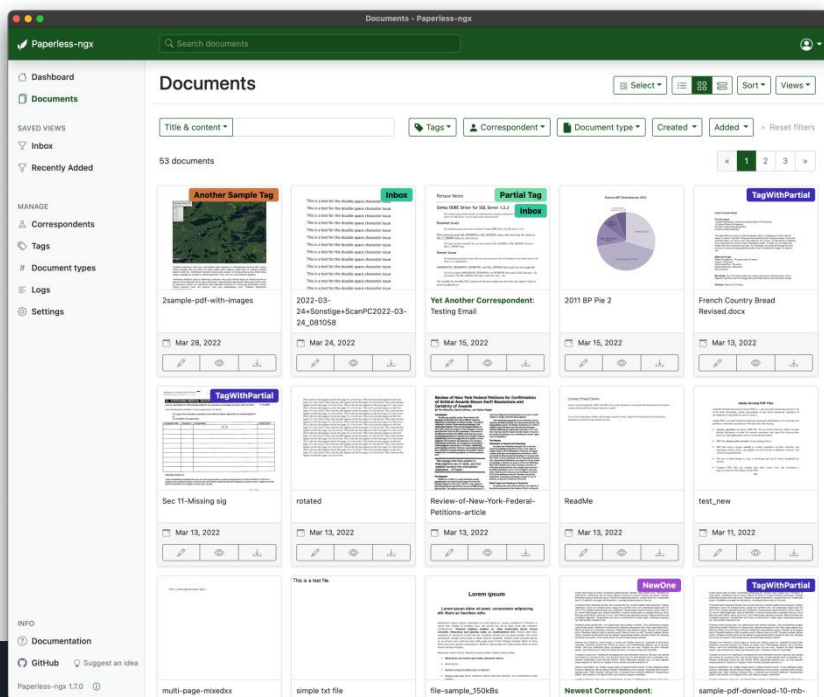
自身影响力

辐射影响力



# 今日热门项目

## paperless-ngx/paperless-ngx



### 项目介绍

Paperless-ngx 是一个社区支持的开源文档管理系统，可将你的物理文档转换为可搜索的在线档案，从而减少纸张的使用。该软件旨在使“无纸化”变得更容易。

### 推荐理由

Paperless-ngx 是一款强大且灵活的开源文档管理系统，自动化处理、多样化的文档支持、OCR 技术和全文搜索功能使其成为优秀的选择。

Stars

15.6k

Forks

785

Watching

91

# 每日GitHub精选

每日 Github 精选将为您展示来自全球开发者的创意与创新，我们精选的项目不仅具有实用价值，更体现了开源社区的活力与智慧。无论是寻找灵感、学习新技术，还是了解行业动态，这个栏目都能为您提供丰富的资源。

Rank	Language	Stars	Forks	Stars Today
01	charlax/professional-programming A collection of learning resources for curious software engineers			
	Python	39849	3061	2808
02	chatchat-space/Langchain-Chatchat Langchain-Chatchat (原Langchain-ChatGLM) 基于Langchain与ChatGLM等语言模型的本地知识库问答			
	Python	22786	3962	181
03	GopeedLab/gopeed A modern download manager that supports all platforms. Built with Golang and Flutter.			
	Dart	9800	752	86
04	krahets/hello-algo 《Hello 算法》：动画图解、一键运行的数据结构与算法教程			
	Java	63392	7712	317
05	1Panel-dev/1Panel 现代化、开源的 Linux 服务器运维管理面板。			
	Go	15885	1422	65
06	Kalabasa/htzm html with targeted manipulation zones			
	JavaScript	1197	25	434
07	zhayujie/chatgpt-on-wechat 基于大模型搭建的微信聊天机器人，同时支持微信、企业微信、公众号、飞书、钉钉接入，可选择 GPT3.5/GPT4.0/Claude/文心一言/讯飞星火/通义千问/Gemini/GLM-4/LinkAI。			
	Python	21884	6063	70
08	FujiwaraChoki/MoneyPrinterV2 Automate the process of making money online.			
	Python	758	89	109
09	ChatGPTNextWeb/ChatGPT-Next-Web A cross-platform ChatGPT/Gemini UI (Web / PWA / Linux / Win / MacOS)			
	TypeScript	61709	51501	353
10	karpathy/minbpe Minimal, clean, code for the Byte Pair Encoding (BPE) algorithm commonly used in LLM tokenization.			
	Python	6488	492	1258

# 每日Gitee精选

每日Gitee精选将为您展示来自国内开发者的创意与创新，我们精选的项目不仅具有实用价值，更体现了开源社区的活力与智慧。无论是寻找灵感、学习新技术，还是了解行业动态，这个栏目都能为您提供丰富的资源。

Rank	Language	Stars	Forks
01	美乐低代码平台 / MetaLowCode 企业级全栈低代码平台 企业级全栈低代码开发平台！ 潜心打造强大核心技术底座： 内置元数据引擎、业务流程引擎、业务触发器、可视化表单、仪表盘设计、智能报表引擎、数据挖掘和插件扩展框架		
	JavaScript	2.1k	511
02	MarsTech/Mars-Admin 管理后台脚手架： Spring Cloud微服务 + Spring Security SSO + JWT Token +自定义登陆页面 + 前后端分离 + Ant Design Pro + K8s部署 +Swagger + 多数据源 + Redis		
	Java	4.6k	446
03	LvHuaiSheng / gzm-design 一款漂亮易用的在线设计器，支持PSD导入、PSD解析，可用于海报设计器、广告设计器、logo设计器、AI创作图片合成器等		
	TypeScript	88	44
04	大漠孤烟/my-tv 我的电视 电视直播软件，安装即可使用		
	C	366	98
05	AntdUI/AntdUI 基于 Ant Design 设计语言的 Winform 界面库		
	C#	368	82
06	TEO Cloud/teo Next-generation web framework for Rust, Node.js and Python.		
	Rust	14	
07	网易有道 / QAnything QAnything (Question and Answer based on Anything) 是致力于支持任意格式文件或数据库的本地知识库问答系统，可断网安装使用		
	Python	104	14
08	WuKongDev / WuKongIM 高性能通用通讯服务，支持即时通讯（聊天软件）(IM)(Chat)，消息推送，消息中台，物联网通讯，音视频信令，直播弹幕，客服系统，AI通讯，即时社区等场景		
	Go	335	100



# 开源之声

在开源的世界里，每一个声音都值得被聆听。为了更好地传递和汇聚开源社区的多元观点，我们特别推出了“每日开源之声”栏目。这个栏目的宗旨是捕捉和分享开源社区内关于最新动态、内容、产品、新闻等话题的用户发声。

“每日开源之声”将每日精选来自国内各大社交平台的热点讨论、用户评论和专业分析，为您提供第一手的开源情报和深度的观点碰撞。我们关注的不仅是技术本身，更关注技术背后的思想火花和社会影响。

## 用户观点

### 程序员因 bug 事故被公司强制要求归还 4 万年终奖

观点1：关键是什么业务？10个人，2023年，一年，赚2000多万，泼天的富贵啊

观点2：可以开除，但是发出去的奖金应该不可能追回来了吧

### 我国 5G 基站总数超 337 万个，5G 移动电话用户达 8.05 亿户

观点1：吓得我看了一眼手机5G开关关了没

观点2：我直接就是关的，太费流量了

观点3：我不是5G，每月2块钱10M的流量套餐

观点4：我长期开5G，绑定移动宽带，每月58，还有特邀5G抽奖，我在外地流量都用不完

### 复盘GitLab史上最严重数据库故障

观点：这么多备份没有一个能用也是离了大谱，平时都没人检查下吗。备份成功与否总该发个邮件

或者通知吧，出问题早该发现了，只能说真是草台班子。

# 媒体观点

## 封面新闻

### 史玉柱将 AI 提升为巨人网络核心业务

22024年2月22日，巨人网络董事长史玉柱在公司内部会议上提出，人工智能（AI）在公司已成为与游戏业务同等重要的新领域。

他预见今年将成为AI技术快速发展的关键时期，表示巨人网络已在这一领域取得不俗成绩，并期待进一步把握机遇。

## 虎嗅网

### 谷歌发开源模型，不 Open 的只剩 OpenAI

一直以来，AI大模型开发和模改的主流生态都被LLaMA占据，直到Mistral出现才略有改观。此番谷歌放出开源Gemma，给开发者提供了更多选择和灵活性，注定会对开源生态起到巨大的刺激作用，促进了开源大模型技术的发展和应创新。

## 量子位

### 英伟达日进5.7亿，黄院士躺印钞机上

黄院士一句话总结：加速计算和生成式AI已达引爆点。

英伟达最新财报出炉。

连创“三高”：

2024财年Q4季度营收达221亿美元（净利122亿美元），比上一季度增长22%，比上一年增长265%。

## 附：中国基础软件开源产业主要参与者

### 基础软件开源项目

#### AI框架



#### 大模型



#### 中间件



## 数据库



## 操作系统



## 开源组织

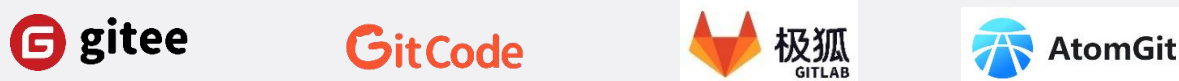
### 开源社区评估机构



### 开源技术论坛



### 开源代码托管平台



### 开源基金会



### 开源产业联盟



# 报告声明

## 分析法说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。

## 社区声明

本报告团队使用合法合规的信息，独立、客观地出具本报告，并对本报告的内容和观点负责。所有编辑人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的意见或观点有直接或间接的联系。

## 免责声明

《开源日报》是开源中国技术社区编辑人员对开源内容相关信息和数据的综合量化评估分析，其性质是为单位或个人研究与观察开源领域发展提供辅助，并不构成最终决策建议。材料结论、数据的完整性、适用性，不予保证，请单位和个人斟酌参考。

## 版权声明

《开源日报》为开源中国技术社区所作，未经开源中国技术社区的审核、确认及书面授权，不得以任何方式在任何媒体上（包括互联网）公开引用本报告的数据和观点，不得以任何方式将本报告的内容提供给其他单位或个人，否则引起的一切法律后果自行承担，同时开源中国技术社区团队亦认为其行为侵犯了开源中国社区团队的著作权，开源中国技术社区团队有权依法追究其法律责任。

# 关键数据来源

所有数据来源于公开数据，包括但不止于媒体门户、技术社区、问答平台、技术论坛、微博、博客、知乎、Gitee、Github、抖音、小红书、B 站、微信公众号、报刊等渠道。