

# 开源日报

OPENSOURCE DAILY



Issue: 26.01.2024  
Date: 29.02.2024



总第019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>
欧盟监管机构拟审查微软对 Mistral 的投资	5
阿里发布肖像视频生成框架 EMO	5
字节打造MegaScale：用于训练LLM的单一 “万卡集群”	6
Meta 计划 7 月发布 Llama 3	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>
库克：苹果今年将在 GenAI 领域 “开辟新天地”	9
Meta 计划 7 月发布 Llama 3	10
<b>今日热门项目</b>	<b>11</b>
<a href="#">olive-editor/olive</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28日 18:00:00~2024年02月29日 18:00:00, 共监测到信息总量7088条, 其中网页1702条 (占比24.01%), APP1512条 (占比21.33%), 微信1208条 (占比17.05%), 微博706条 (占比9.96%), 视频680条 (占比9.59%), 问答560条 (占比7.9%), 头条号367条 (占比5.18%), 论坛343条 (占比4.84%), 报刊10条 (占比0.14%)。



### 新闻平台

数据: 1702条 占比: 24.01%



### 社交平台

数据: 1914条 占比: 27.01%



### 视频平台

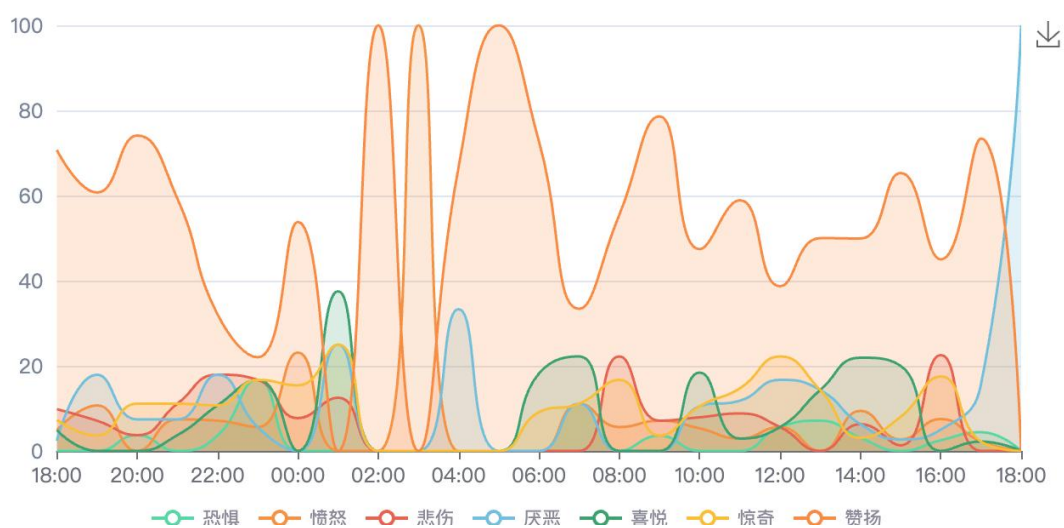
数据: 680条 占比: 9.59%



### APP平台

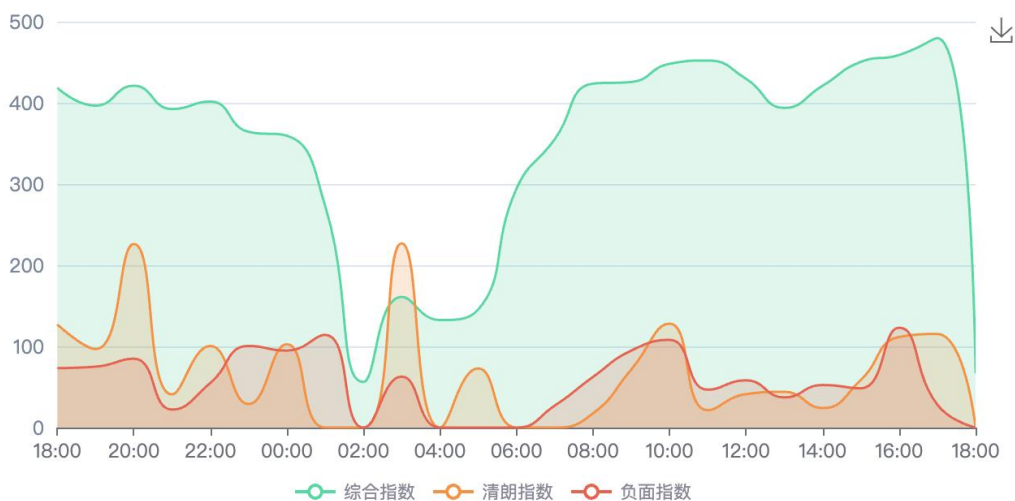
数据: 1512条 占比: 21.33%

## 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



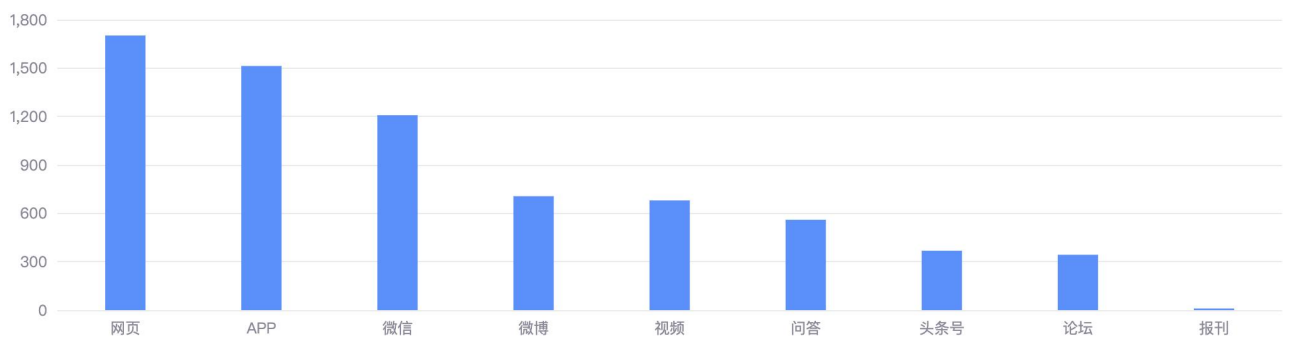
舆情指数模型POI，根据舆情系统目前可采集数据设定。针对正面舆情跨平台传播热度进行评估（网页、微信、微博、客户端文章数量和正面舆情数量占比），暂时未考虑媒体影响力、事件敏感程度、历史风险性等其他要素。

## 开源话题热词走势



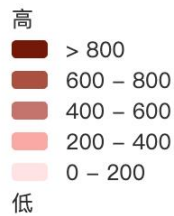
今日开源话题热词中，“数据”“图像”“EMO”等关键词在社交平台中热度较高

## 渠道传播TOP



分析时段内，网页信息1702条（占比24.01%）、APP 信息1513条（占比21.34%）、微信信息1208条（占比17.04%）、微博信息706条（占比9.96%）、视频信息680条（占比9.59%）、问答信息560条（占比7.89%）、头条号信息368条（占比5.19%）、论坛信息343条（占比4.84%）、报刊信息10条（占比0.14%）。

地区	数量
北京	963
广东	308
上海	238
山东	53
浙江	52
福建	48
河南	47
四川	36
江苏	35
安徽	34

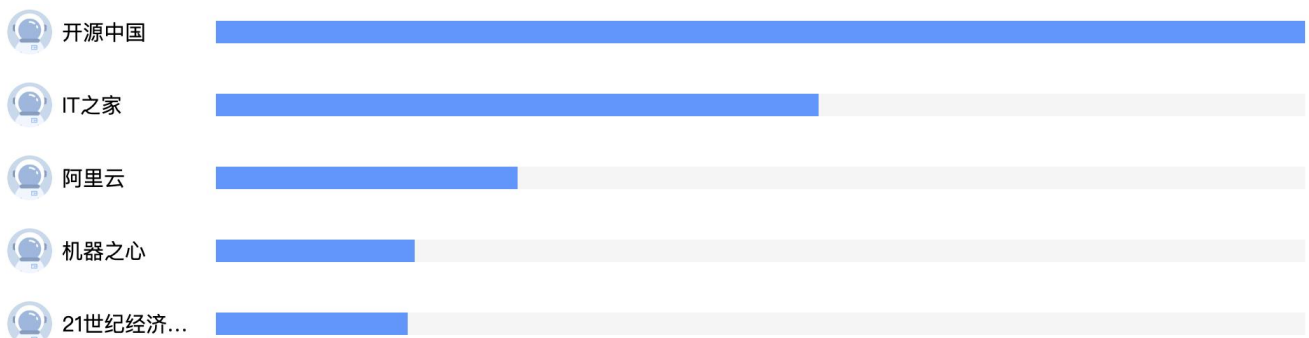


## 开源话题区域提及走势

分析时段内，发布信息地域信息地域主要集中在北京963条（占比48.59%）、广东308条（占比15.54%）、上海238条（占比12.01%）、山东53条（占比2.67%）、浙江52条（占比2.62%）。

今日北京、广东、上海地区对开源话题的兴趣度占比在 75% 左右。

## 今日活跃度渠道评估



# 今日要点 | TODAY FOCUS

## 阿里发布肖像视频生成框架 EMO

来自阿里巴巴的团队发布了音频驱动的肖像视频生成框架EMO (Emote Portrait Alive)，相关论文同步发表于arXiv。输入一张参考图像和声音音频，该框架能够生成具有丰富面部表情和头部姿势的声音肖像视频。

### 编辑简评

阿里发布的EMO框架在人像视频生成领域是一项重大的技术突破，它为视频内容创作和虚拟角色动画制作开辟了新的可能性。

这项技术不仅在娱乐产业中具有广泛的应用潜力，还可以用于学术研究和教育培训等领域。然而，需要注意的是，该框架目前主要用于学术研究和效果演示，未来还需要进一步优化和扩展其应用范围。

## 欧盟监管机构拟审查微软对 Mistral 的投资

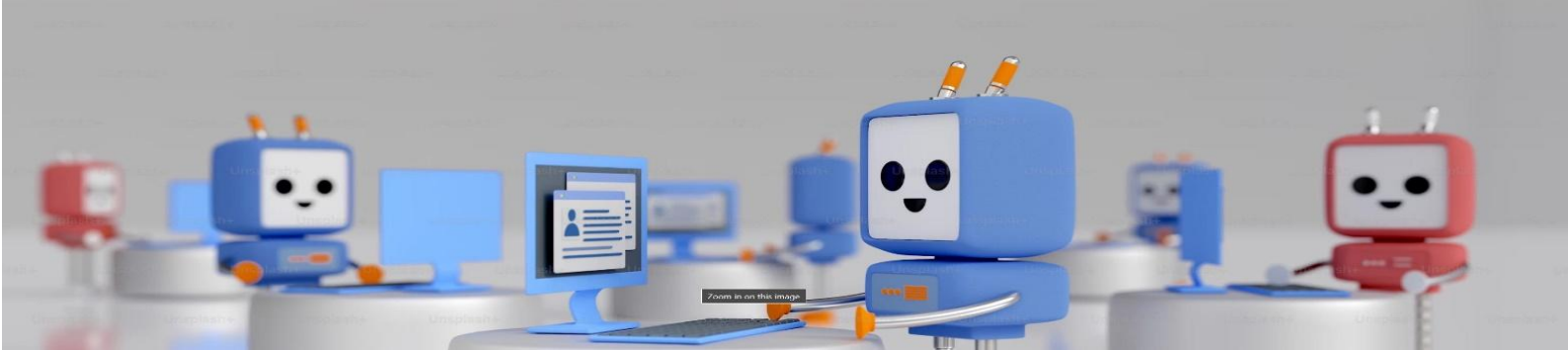
据The Information报道，欧洲最高反垄断监管机构欧盟委员会当地时间周二称，计划审查微软对法国生成式AI初创公司Mistral的投资。微软于周一宣布与Mistral建立合作伙伴关系，向Mistral投资1630万美元，并在云平台Azure中提供Mistral的模型。

### 编辑简评

Mistral AI最新研发的语言模型“Mistral Large”将被整合至微软的Azure云计算平台，提供先进的人工智能服务。Mistral AI的模型在多项常用基准测试中表现卓越，仅次于OpenAI的GPT-4，成为全球排名第二的先进模型。

微软此次投资被视为其在寻求多元化人工智能合作伙伴战略中的又一重要举措。此前，微软已向OpenAI投资了数十亿美元。微软与OpenAI的紧密联盟已经在美国和欧盟引发了潜在的监管审查，因此，微软对Mistral AI的投资无疑将进一步加剧监管机构对于科技巨头在人工智能领域垄断行为的担忧。





## Meta 计划 7 月发布 Llama 3

根据 The Information 的报道，Meta 计划在 7 月发布最新版本的大语言模型 —— Llama 3。Meta 的一位员工透露，由于研究人员尚未开始对模型进行微调，因此公司尚未决定 Llama 3 是否是多模态的。

Meta 员工还说，Llama 3 最高可能拥有超过 1400 亿个参数，相比之下，去年 7 月发布的 Llama 2 的参数最高为 700 亿。

### 编辑简评

Meta 在去年 7 月发布的 Llama 2 中引入了保护措施，以防止模型回答被认为有争议的问题。然而，这些措施在某些情况下可能显得过于保守，例如在拒绝回答如何恶作剧朋友或赢得战争等争议性不强的问题时。因此，Meta 的研究人员正在尝试放松 Llama 3 的保护措施，以便它能更自然地与用户互动。

Llama 3 的开发反映了人工智能公司面临的挑战：在创造有趣、互动的产品的同时，避免因不恰当或错误的回答而引起公关灾难。此外，Llama 3 作为开源模型，任何人都可以免费使用，这也带来了关于如何处理人工智能安全性的讨论。

## 字节打造MegaScale：用于训练LLM的单一“万卡集群”

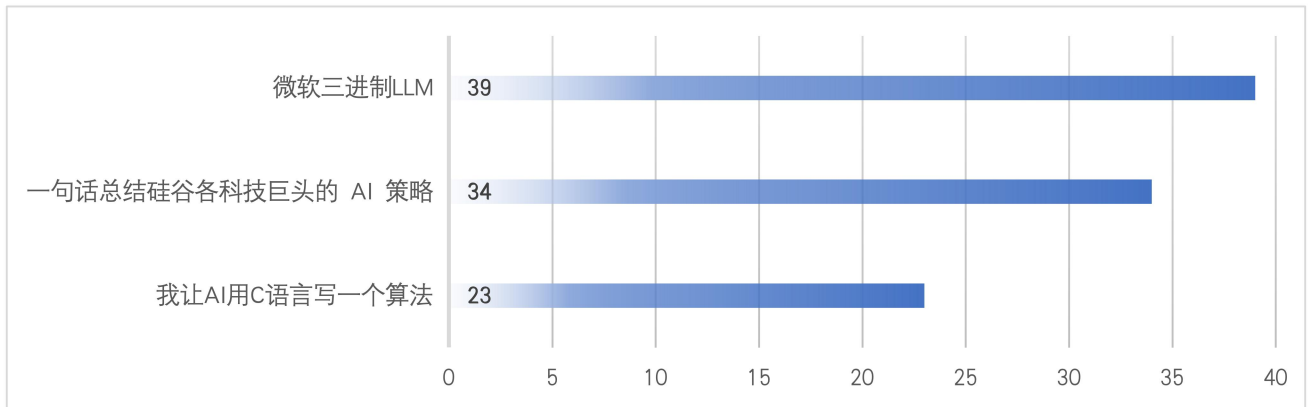
来自字节和北大的团队在arXiv上发表论文，提出大模型生产系统MegaScale。MegaScale将大型语言模型训练扩展到超过10000个GPU，在12288个GPU上训练175B LLM模型时，MegaScale实现了55.2%的模型FLOP利用率（MFU），与层内模型并行技术Megatron-LM相比，MFU提高了1.34倍。

### 编辑简评

MegaScale作为字节跳动在人工智能领域的一个重要项目，不仅展示了其在处理大规模计算任务方面的能力，也为未来LLM的发展和应用提供了强大的技术支持。

通过这种大规模集群的应用，可以预期未来会有更多创新和高效的人工智能模型出现，推动整个行业的进步。

## 社交观察



### 我让 AI 用 C 语言写一个算法

我让AI用C语言写一个算法，它写了一个调用Python虚拟机的C代码，然后在Python虚拟机里传了一段Python的实现，这歪门邪道都是哪儿学来的( ; ´Д` )

- 微博 一个动态类型的幽灵

### 一句话总结硅谷各科技巨头的 AI 策略

OpenAI: 大力出奇迹的 AI; Meta: 开源的 AI; 亚马逊: 云端的 AI; 英伟达: 硬件的 AI; 特斯拉: 路上跑的 AI; 谷歌: 每个组都要出 AI; 微软: 捡漏 OpenAI; 苹果: 赶紧不造车了造 AI

你们觉得哪家最有潜力？。

- 微博 甜菜欣欣

### 微软三进制 LLM

现在，大语言模型（LLM）迎来了“1-bit时代”。这就是由微软和中国科学院大学在最新一项研究中所提出的结论——所有的LLM，都将是1.58 bit的。

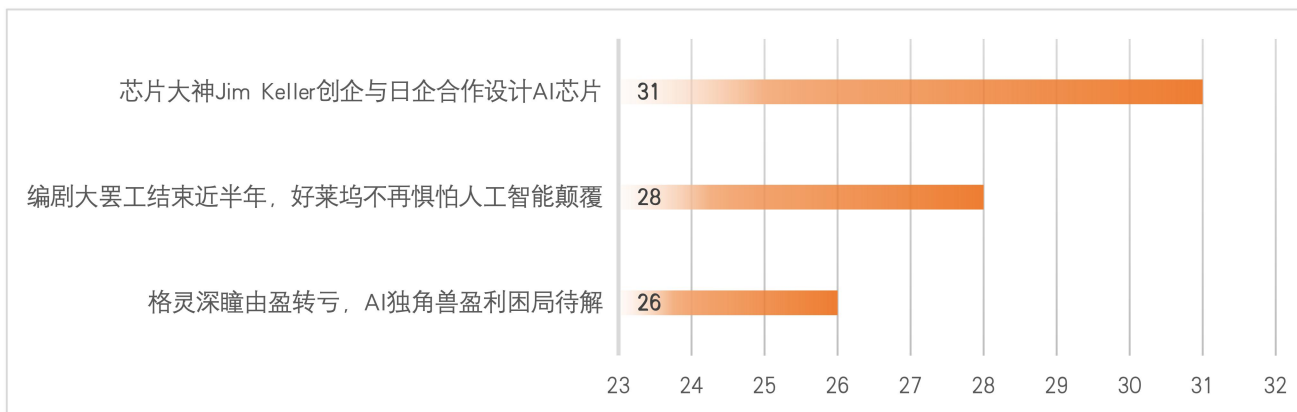
具体而言，这项研究提出的方法叫做BitNet b1.58，可以说是从大语言模型“根儿”上的参数下手。

将传统以16位浮点数（如FP16或BF16）形式的存储，统统变成了三进制，也就是  $\{-1, 0, 1\}$ 。

值得注意的是，这里的“1.58 bit”并不是指每个参数占用1.58字节的存储空间，而是指每个参数可以用1.58位的信息来表示。

- 微博 量子位

媒体观察



编剧大罢工结束近半年，好莱坞不再惧怕人工智能颠覆

这场持续了148天、令整个行业陷入停滞的好莱坞编剧大罢工于去年9月底正式结束，劳资双方达成协议。在备受关注的AI问题上，协议并未完全禁止AI的使用，而是对多个具体情境进行了规定，以保证编剧不会沦为AI的修改工具：比如AI生成的素材不能被视为原始素材，编剧的署名权和独创性不可因AI的介入而受损。制片公司不得要求编剧使用AI，编剧工会有权禁止制片公司将编剧创作的素材用于训练AI。

- 界面新闻

芯片大神 Jim Keller 创企与日企合作设计 AI 芯片

日本政府支持的半导体开发研究小组（前沿半导体技术中心或LSTC）将与美国初创公司Tenstorrent合作设计其首款先进人工智能（AI）芯片。

Tenstorrent由业内资深人士Jim Keller领导，公司采用开源RISC-V标准，旨在为客户提供一种替代领先者英伟达和Arm的选择。除了Keller之外，Tenstorrent还聘请了在AMD工作了13年的Keith Witek担任首席运营官，以及Wei-han Lien担任首席芯片架构师。

- 爱集微

格灵深瞳由盈转亏，AI 独角兽盈利困局待解

格灵深瞳作为机器视觉赛道上的玩家之一，进入2023年，营收下降，并由盈转亏。格灵深瞳在业绩快报中指出，受宏观经济环境、行业终端客户预算及采购计划推迟等多重因素影响，公司目标客户需求滞后，交付需求延迟，导致营业收入较上年同期出现一定程度的下滑。

整体来说，人工智能公司商业变现还是比较困难，而持续亏损的AI独角兽，AI落地之路更是步履维艰。

- 网易

# 事件点评

## 库克：苹果今年将在 GenAI 领域 “开辟新天地”



苹果公司首席执行官蒂姆 - 库克 (Tim Cook) 在公司年度股东大会上承诺称, 苹果今年将在 GenAI 方面 "开辟新天地"。就在前几日, 苹果公司已经宣布取消了其耗资数十亿美元、长达十年的电动汽车制造计划, 而电动汽车项目的部分员工则被重新分配到了 GenAI 的各项计划中。

彭博社报道称, 苹果公司正计划利用 GenAI 模型升级 Siri 和 iOS 内置搜索工具 Spotlight, 目的是让它们能够回答更复杂的查询, 并处理复杂的多轮对话。据称, 苹果还在探索 AI 功能, 让用户可以在 Keynote 中自动生成演示幻灯片, 在 Apple Music 中自动生成播放列表, 并在公司的应用程序开发平台 Xcode 中提供由 GenAI 支持的编码建议。



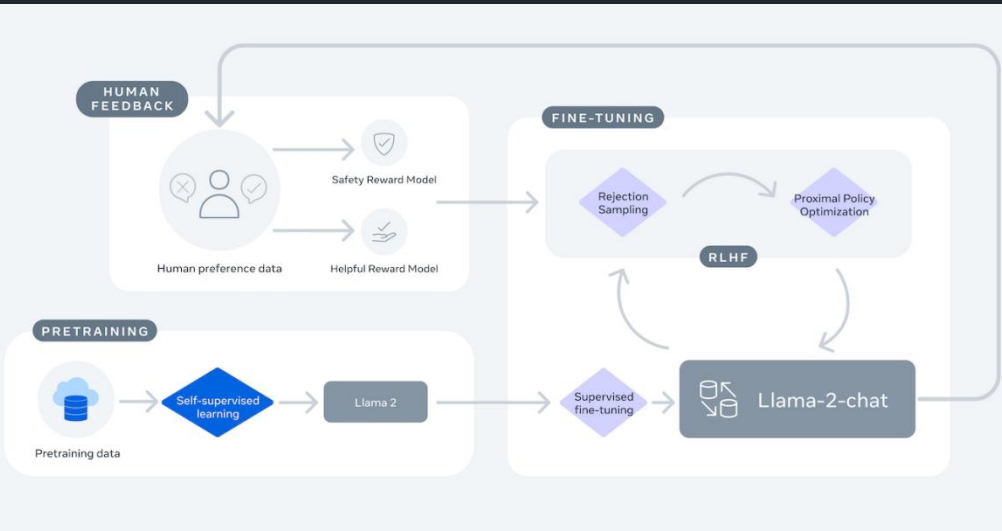
事件点评 Event Search

苹果在AI领域的这一举措, 可能会对整个行业产生深远影响。它不仅可能推动AI技术的进一步发展和应用, 还可能对竞争对手的策略产生影响, 促使整个行业在AI技术上投入更多资源。

自身影响力

辐射影响力





# Meta 计划 7 月发布 Llama 3

根据 The Information 的报道，Meta 计划在 7 月发布最新版本的大语言模型 —— Llama 3。Meta 的一位员工透露，由于研究人员尚未开始对模型进行微调，因此公司尚未决定 Llama 3 是否是多模态的。Meta 员工还说，Llama 3 最高可能拥有超过 1400 亿个参数，相比之下，去年 7 月发布的 Llama 2 的参数最高为 700 亿。参数数量越多，通常意味着模型能提供的结果越精确，不过较小的模型可能在处理速度上更有优势。Llama 3 正在由 Meta 旗下的独立基础人工智能研究团队（FAIR）以外的生成人工智能团队开发。



## 事件点评 Event Search

Meta的这些举措显示了其在AI领域的野心，尤其是在开源生成式AI领域。Meta不仅希望打破OpenAI在大语言模型市场的主导地位，还希望通过开源社区模型的快速开发从中受益。

总的来说，Llama 3的推出预计将对AI行业产生重大影响，特别是在推动开源技术发展和与闭源模型竞争方面。

但Meta仍需克服人才流失等问题，以确保Llama 3能够持续发展并产生广泛影响。

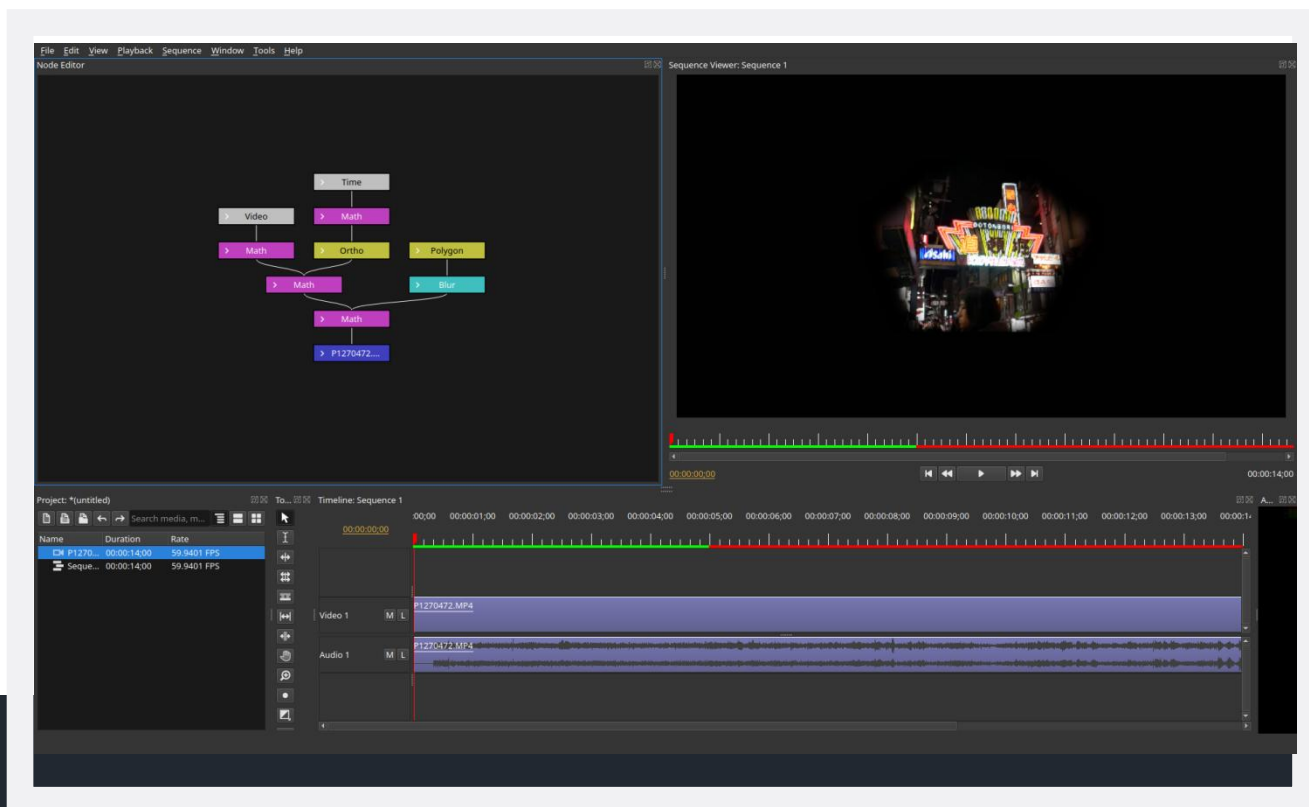
自身影响力

辐射影响力



# 今日热门项目

olive-editor/olive



## 项目介绍

Olive 是跨平台的开源非线性视频剪辑工具，支持 Windows、macOS 和 Linux，目标是提供高端专业视频编辑软件的免费替代方案。

## 推荐理由

在 Linux 平台上，专业的视频编辑软件选择相对有限，尤其是免费且开源的软件。Olive 的出现填补了这一市场空缺，为 Linux 用户提供了更多选择。Olive 是一款适合喜欢探索新工具、愿意为开源项目做出贡献的程序员尝试的视频编辑软件。虽然它还在不断发展中，但其潜力和已经实现的功能使它成为一个值得关注的选择。

Stars

7.7k

Forks

522

Watching

174

# 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	Lissy93/web-check All-in-one OSINT tool for analysing any website			
	TypeScript	9635	750	257
02	pmndrs/uikit user interfaces for react-three-fiber			
	TypeScript	1082	51	80
03	OpenCodeInterpreter/OpenCodeInterpreter OpenCodeInterpreter is a suite of open-source code generation systems aimed at bridging the gap between large language models and sophisticated proprietary systems like the GPT-4 Code Interpreter.			
	Python	937	135	62
04	movie-web/movie-web A small web app for watching movies and shows easily			
	TypeScript	19056	2057	58
05	toeverything/AFFiNE AFFiNE(pronounced [ə 'fain]) is a next-gen knowledge base that brings planning, sorting and creating all together. Privacy first, open-source, customizable and ready to use.			
	TypeScript	27699	1788	403
06	joaomdmoura/crewAI Framework for orchestrating role-playing, autonomous AI agents. By fostering collaborative intelligence, CrewAI empowers agents to work together seamlessly, tackling complex tasks.			
	Python	8540	953	86
07	public-apis/public-apis A collective list of free APIs			
	Python	283281	30924	1709
08	tigerbeetle/tigerbeetle The distributed financial transactions database designed for mission critical safety and performance.			
	Zig	5482	272	563
09	taikoxyz/taiko-mono A decentralized, Ethereum-equivalent ZK-Rollup.			
	HTML	3274	1550	198
10	dockur/windows Windows in a Docker container.			
	Shell	2277	128	283

# 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	dromara/Sa-Token 一个轻量级 Java 权限认证框架，让鉴权变得简单、优雅！—— 登录认证、权限认证、分布式Session会话、微服务网关鉴权、单点登录、OAuth2.0		
	Java	28k	3.3k
02	RockYang/chatgpt-plus ChatGPT-PLUS 基于 AI 大语言模型 API 实现的 AI 助手全套开源解决方案，自带运营管理后台，开箱即用。集成了 OpenAI, Azure, ChatGLM,讯飞星火，文心一言等多个平台的大语言模型。采用 Go + Vue3 + element-plus 实现。		
	Go	1k	457
03	panday/chat-master-web 一个已对接市面上主流大模型ChatGPT(3.5、4.0)模型、文心一言、通义千问、讯飞星火、智谱清言(ChatGLM)的平台。支持一键切换模型、支持文心一言(支持Stable-Diffusion-XL作图)、支持联想上下文对话。		
	JavaScriptt	53	25
04	fufei/fly-barrage 弹幕库 功能完善，强大的 web 端弹幕库（Fully functional and powerful web-based barrage library）。包含完整 DEMO，可快速上手。		
	TypeScrip	62	8
05	Token/FastWiki 本项目是一个高性能、基于最新技术栈的知识库系统，专为大规模信息检索和智能搜索设计。		
	C#	30	9
06	tujiaw/ntsscreenshot Windows截图工具		
	C++	46	22
07	LuXu/oh-my-foss-android (镜像仓库) 个人收集的实用、良心开源安卓软件		
		110	15
08	CTO说/系统分析师 2024年系统分析师（软考高级）备考资源库。		
	HTML	188	126



# 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

### 替代 Nginx, Cloudflare 开源 Pingora Rust 框架

观点：打卡，终于来了，不知道这货，能不能带给 10 多年前，用 Nginx 替换掉公司 F5 的绝大部分功能时候那种激动！！

### 某公司拿「织梦CMS」“碰瓷”失败，法院判GPL开源协议有效

观点1：干得漂亮，如果有什么公益组织能代理这些开源软件在国内起诉这吸血公司就更好了

判决：一审法院认为 DedeCMSV5.7-sp1 中包含 GPL 协议下开源的代码，整体应遵守 GPL 协议，A 医院使用软件不用支付授权费用，但仍需遵守署名权的要求。

观点2：GPL真是神来之笔，承认GPL协议在中国的法律效力，预示着中国的开源事业与世界开源接轨，但法院对GPL条款的解读还有待提升！GPL是用来保障用户的自由,却被法院错误的判成了如何让用户变成不自由!无锡中院把GPL3.0第9条好好看看吧~离事实仅差一步之遥！

观点3：给无锡中院点赞，给想以此碰瓷的人当头一棒，gpl许可协议不是你想咋解读就咋解读

### 库克：苹果今年将在 GenAI 领域 “开辟新天地”

观点：没戏，这不是苹果擅长的事，如果是造车的话还可以期待下

# 媒体观点

## 证券日报

### Sora 为 AI 带来更多可能性 企业亟须突破算力瓶颈

在经历了去年AI大模型爆发式增长后，AI技术迎来新一轮更新迭代。文生视频模型Sora的出现，开启了AI的新时代，也对企业的算力提出更高的要求。

浙江大学国际联合商学院数字经济与金融创新研究中心联席主任、研究员盘和林对《证券日报》记者表示，Sora的出现，让国内AI企业趋之若鹜，但实现Sora同等级应用的难度并不小，对算力的需求会越来越大，所以各大AI企业应从现实出发，考虑应用落地的难易度。

## 21世纪经济报道

### 荣耀猛攻！挑战苹果三星

近一年来，荣耀在AI赛道上凶猛下注，尤其从操作系统、人机交互层面进行升级。在MWC上，荣耀正式发布了全新的AI使能的全场景战略，推出平台级AI赋能，以人为中心的跨操作系统体验和基于意图识别的全新人机交互，并发布了旗下首款AI PC。

在AI策略的基础之上，荣耀进一步在产品上、高端市场上向三星和苹果发起挑战，向世界荣耀的目标加速迈进。

## 光子星球

### AI 硬件，距离 iPhone 时刻仅一步之遥？

当 AI 与硬件实现耦合的场景愈发丰富，AI 终于开始摆脱生产力工具的固有印象，其技术落地的场景愈发贴近消费语境。如曾在去年为业内热议的 Humane 的 AI Pin 和 Rabbit 的 R1，均试图跳脱出既往智能终端的束缚，从而更贴近我们的生活。正因如此，此时我们更应该关注这些未被商业竞争所桎梏的创新，兴许能为国内大模型的商业落地提供更多视角。

## 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

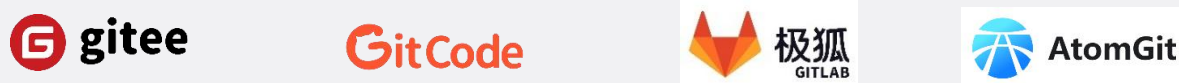
### 开源社区评估机构



### 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，包括但不限于媒体门户、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。