



# 开源技术态势数据观察 大模型开发生态全景、趋势和观点

2025.05

演讲人：王旭 蚂蚁开源技术委员会副主席 —

蚂蚁的开源团队其实不叫开源办公室（OSPO），而是叫“**开源技术增长**”，是蚂蚁的一个技术架构性团队，我们的一个工作目标就是“利用对开源社区的洞察来为蚂蚁的架构和技术的演进提供指引”，我们做的这样的全景图和趋势分析，也正是希望能依据对社区的洞察，指出哪些项目是最应该跟踪、使用、参与的，哪些方向是值得投入的，反之亦然。

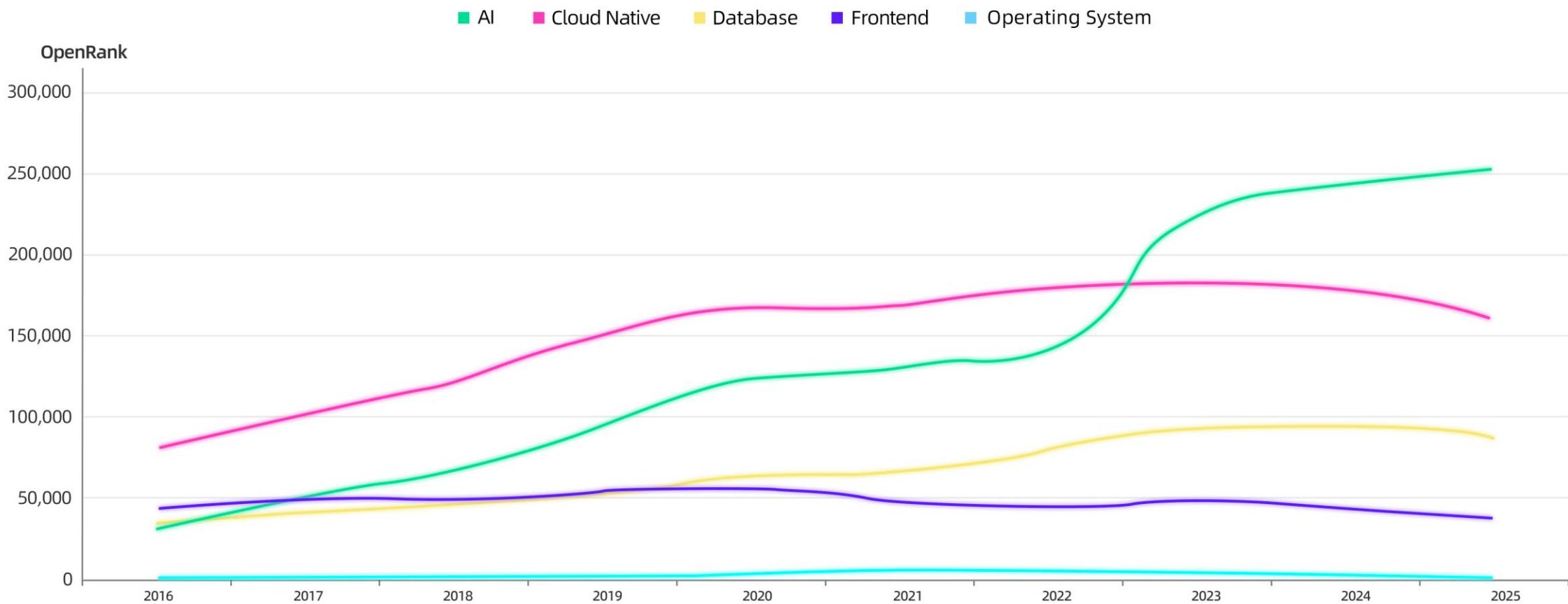


王旭 蚂蚁开源技术委员会副主席



根据 OpenDigger 的数据显示，

「人工智能已经于 2023 年超越云原生，成为协作影响力最大的技术领域」



## 如何做一张生态全景图

我们选取了时下 AI 领域中最耳熟能详的项目（例如 PyTorch, LangChain, vLLM 等）作为种子节点，通过开发者在 GitHub 上的不同项目之间产生的协作关联关系，不断去探查生态的多个侧面。在得到初始的项目列表的基础上，通过和不同的专家探讨，手动的维护项目所属的技术领域标签，并不断做列表的收敛和更新。

AI 技术迭代一日千里，我们希望呈现的是当下最顶尖的和最火热的开源项目，因此，使用了华东师范大学 X-lab 实验室的 OpenRank 影响力评价指标作为一个重要的数据依据：在今年的 OpenRank 月均值大于 10 的项目才会出现在全景图上，一些年久失修，或者是过于崭新的项目，如果没有达到这个阈值，则会被拿掉。

最终，我们呈现和发布 **2025 年大模型开源开发生态全景图**，一共包含 135 个项目，涵盖了智能体应用层和模型基础设施层一共 19 个技术领域。

[1] OpenRank 指标：[https://open-digger.cn/docs/user\\_docs/metrics/community\\_openrank](https://open-digger.cn/docs/user_docs/metrics/community_openrank)

# Open Source LLM Development Landscape 2025



Application

General Assistant	<b>OpenManus</b> OWL Second Me	Coding Assistant	<b>bolt.new</b> OpenHands aider	Continue Cody Tabby PR-Agent	Cline	AI Search Engine	Perplexica Scira SearXNG	Client Interface	<b>Open WebUI</b> SillyTavern Text generation web UI	Lobe Chat NextChat Chatbox
Agent Framework	<b>Dify</b> n8n	RAGFlow	FastGPT MaxKB	Flowise AI AUTOapt	DB-GPT LangChain-Chatchat	haystack		Tool Use	ComfyUI Vercel	Browser Use OmniParser
	LangChain LlamaIndex	AutoGen	Semantic Kernel	LangGraph				API Management	<b>LitELLM</b> A2A protocol	Model Context Protocol compositio One API
	crewai Agent Development Kit	OpenAI Agents SDK CAMEL-AI	GraphRAG	mem0 Letta	MetaGPT					

Infrastructure

Model Training, Development and Serving										
Data Transformation			Model Training, Development and Serving					APP Integration		
Data Labeling			Serving					Evaluation Platform		
Data Labeling			Ollama	Xorbits Inference	NVIDIA Dynamo	Triton Inference Server	Evaluation Platform			APP Framework
Data Integration			gpt4all	KServe	Mooncake	AlBrix	BentoML	Streamlit gradio		
Apache Airflow			LLM SGL	OpenVINO	ONNX RUNTIME	LMDeploy	Text Generation Inference	Vector Storage and Search		
dagster			Transformers	TensorRT-LLM	elasticsearch			Milvus		
dbt Core			LLaMA-Factory	unsloth	verl	OPENRLHF	Gymnasium	OpenSearch		
DataChain			Easy and Efficient LLM Fine-Tuning					weaviate		
Data Governance			PyTorch	Paddle	NeMo	Megatron-LM	Weights & Biases	chroma		
ICEBERG			Post-train	TensorFlow	Keras	deepspeed	Colossal-AI	drant		
APACHE Paimon			Distributed Processing	RAY	APACHE Spark	VOLCANO	dask	LanceDB		
APACHE hudi			Format					pgvector		
Open Metadata										
DataHub										
APACHE GRAVITINO										
Unity Catalog										
FlashAttention			AI Kernel Library	CUTLASS	MLX	FlashInfer	AI compiler	Triton	OpenXLA	ZenML
NCCL			DeepEP							

# 2025 大模型开源开发生态 OpenRank Top 20 项目



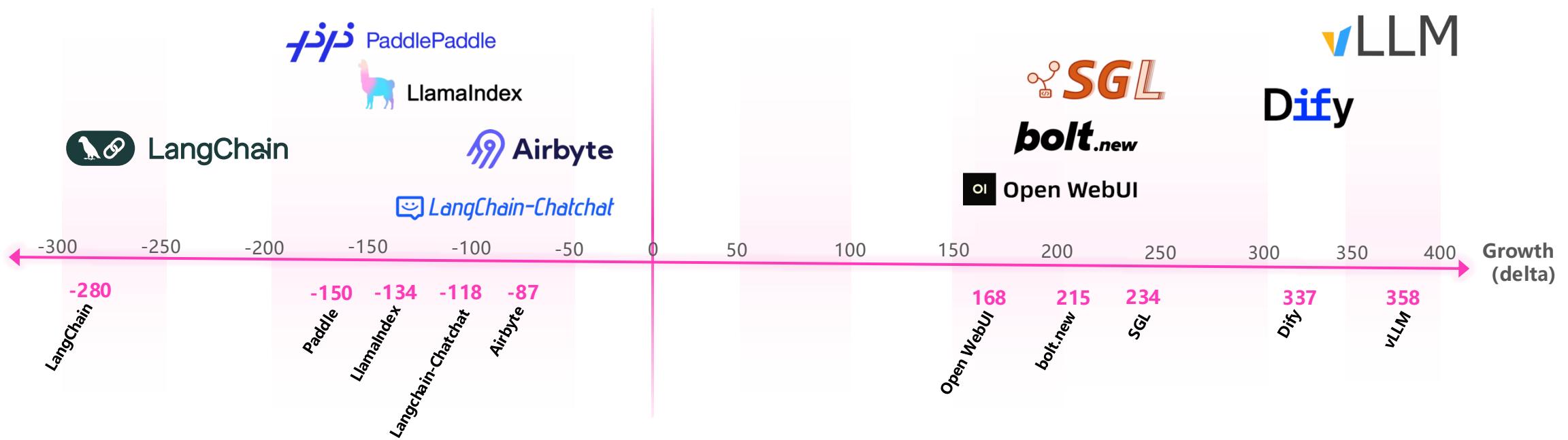
#	项目名	技术领域	OpenRank	Star	社区规模	语言	创建时间	主要发起组织
1	PyTorch	模型训练	927	89,600	2,143	Python	2016-08-13	Meta
2	LLM	模型服务	615	46,409	3,161	Python	2023-02-09	UC Berkeley
3	Dify	应用框架	473	95,414	3,243	TypeScript	2023-04-12	Dify.AI
4	elasticsearch	向量搜索	349	72,542	463	Java	2010-02-08	Elastic
5	Ollama	模型服务	311	139,289	3,109	Go	2023-06-26	-
6	OpenVINO	模型服务	280	8,220	450	C++	2018-10-15	Intel
7	Airflow	数据集成	276	39,906	687	Python	2015-04-13	Airbnb
8	SGL	模型服务	269	13,902	1,107	Python	2024-01-08	UC Berkeley
9	Open WebUI	应用工具	245	92,676	1,867	JavaScript	2023-10-06	-
10	LiteLLM	应用工具	238	21,795	1,233	Python	2023-07-27	-
11	LLaMA	模型服务	232	79,118	895	C++	2023-03-10	ggml.ai
12	RAY	并行计算	227	36,845	571	Python	2016-10-25	Anyscale
13	PaddlePaddle	模型训练	218	22,731	396	C++	2016-08-15	百度
14	bolt.new	编程助手	215	14,503	3,462	TypeScript	2024-09-24	Stackblitz
15	n8n	应用框架	213	88,153	1,252	TypeScript	2019-06-22	n8n
16	RAGFlow	应用框架	208	51,009	1,426	TypeScript	2023-12-12	InfiniFlow
17	Airbyte	数据集成	203	18,032	594	Python	2020-07-27	Airbyte Inc
18	ONNX	模型服务	201	16,451	564	C++	2018-11-10	微软
19	Spark	并行计算	199	41,033	260	Scala	2014-02-25	UC Berkeley
20	Nemo	模型训练	193	13,771	308	Python	2019-08-05	NVIDIA

声明：所有 logo 均为引用，版权归其所在组织所有

注：以上数据截止 2025 年 4 月 30 日

# 2024 vs 2025

## 同比增长最多和降低最多的项目们

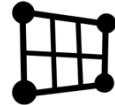


声明：所有 logo 均为引用，版权归其所在组织所有

注：分别取 2024 年和 2025 年 1 月至 4 月的项目 OpenRank 均值，计算同比绝对值变化

# 三大主导的技术赛道

## 01 模型训练框架



PyTorch 是当之无愧的生态顶流，在全景图中的所有项目中影响力位列第一。而国产的深度学习平台百度飞桨，对比去年同期 OpenRank 降低了 41%。

## 02 高效推理引擎



高效推理引擎 vLLM 和 SGLang 在过去一年都处于飞速迭代之中，分别位于 OpenRank 同比增长的第一和第三位，他们以优越的 GPU 推理性能优势在企业级 LLM 部署中广受拥趸。

## 03 低代码应用开发框架



结合低代码工具链和 RAG 知识检索管理技术的 Dify 和 RAGFlow，正在迎接属于他们的高速增长，而这两个应用开发平台都是从中国开发者社区中生长出来的强势项目。

我们看了 100 多个大模型开发领域相关的开源项目，  
然后发现...



A Real-world Hackathon

# 大模型开发生态是一场 真实世界的黑客松

在 AI 技术扩散速度远超预期的背景下，大模型开发生态正演变为一场在真实世界发生的、实时公开直播的黑客松。开发者借助 AI 的力量得以以“超级个体”的姿态，在热点事件中快速构建开源项目，通过技术能力争夺公共流量与行业话语权，与此同时，一波一波项目也在技术概念浪潮中爆火或者消亡。这种快速构建，快速消亡，注重开发速度的迭代，正如闭关数天出项目的黑客松一般。



# 开发者的短线投资与技术声誉

当 Devin、Perplexity、Manus 等闭源产品引发行业震动时，开发者们迅速复刻出[开源版本](#)：



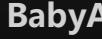
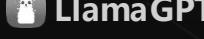
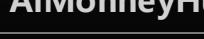
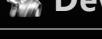
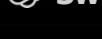
(后改名为： OpenHands)



这些项目的生命力究竟如何，还需要时间验证，但开发者通过技术声誉实现了自身的长期价值 -- GitHub Star 数、社区讨论热度以及与头部机构的合作机会，成为衡量开发者“影响力资产”的新指标。即便项目最终沉入“AI 墓园”，开发者也能凭借早期贡献获得行业认知的红利。

# AI “墓园” - 昙花一现的技术实验

2023 年开始的 LLM 浪潮催生了一批“速生速死”的 AI 开源项目。如 BabyAGI 和 Swarm 这样的项目，在发布时就声称是“实验性”、“探索性”的尝试，也许从一开始就没做深谋远虑的长期规划。然而他们所提出的先锋概念，激发的讨论和创新尝试，在推动这场黑客松接力赛从“概念验证”向“工程落地”演进。

项目	项目描述	Star 数量	创建时间	最后提交时间
 <b>bloop.</b>	用 Rust 编写的代码搜索引擎	9,467	2022-12-10	2024-12-04
 <b>DroomGPT</b>	上传房间照片即可生成梦想房间	10,318	2023-02-21	2024-04-20
 <b>Chatbot UI</b>	在任何模型上 AI 聊天	31,301	2023-03-11	2024-08-03
 <b>Dalai</b>	本地运行 LLaMA 的最简方式	13,082	2023-03-12	2024-06-18
 <b>Mr. Raneddeer</b>	GPT-4 AI 个性化导师 prompt, 支持自定义学习体验	29,533	2023-03-31	2024-03-25
 <b>BabyAGI</b>	旨在开发能够自我构建和自我扩展的自主智能代理	21,476	2023-04-03	2024-11-06
 <b>LlamaGPT</b>	本地自托管 ChatGPT 聊天机器人 (Llama 2 驱动)	10,971	2023-07-22	2024-04-23
 <b>draw-a-ui</b>	画出原型图自动生成 HTML	13,492	2023-11-07	2024-07-18
 <b>AiMonneyHunte</b>	AI 副业赚钱项目集合	14,680	2023-12-14	2024-12-21
 <b>Devika</b>	Agentic AI 软件工程师, 能理解高阶人类指令并写代码	19,257	2024-03-21	2024-09-19
 <b>Swarm</b>	轻量级多智能体调度的教育型框架	19,809	2024-02-22	2025-03-11

声明：所有 logo 均为引用，版权归其所在组织所有

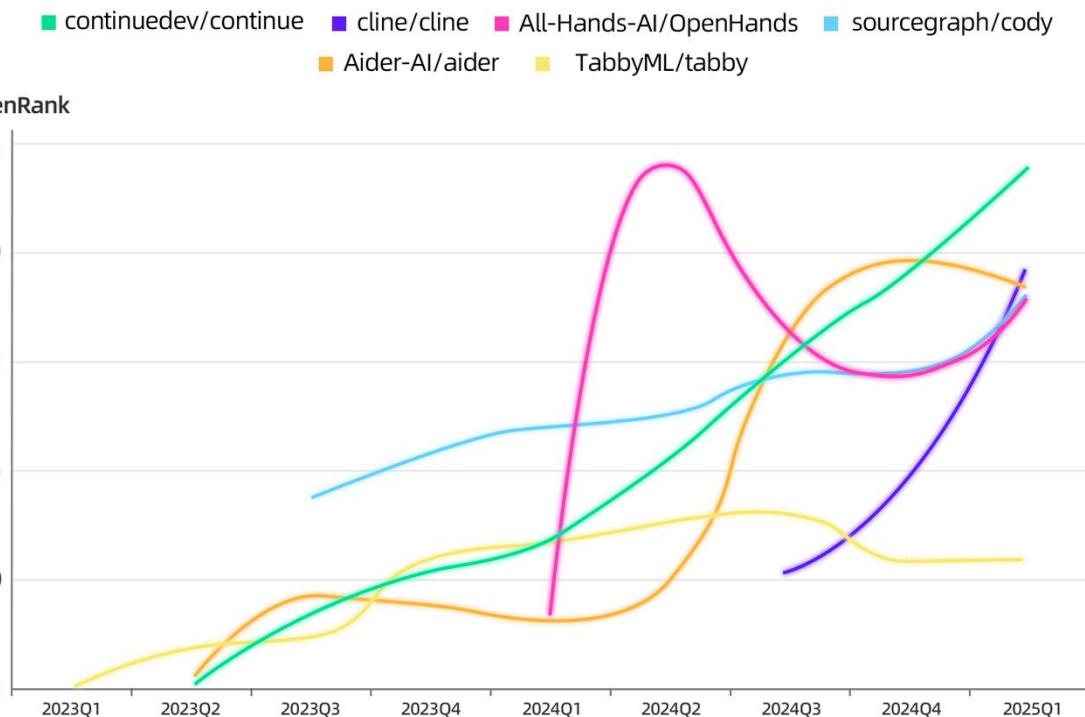
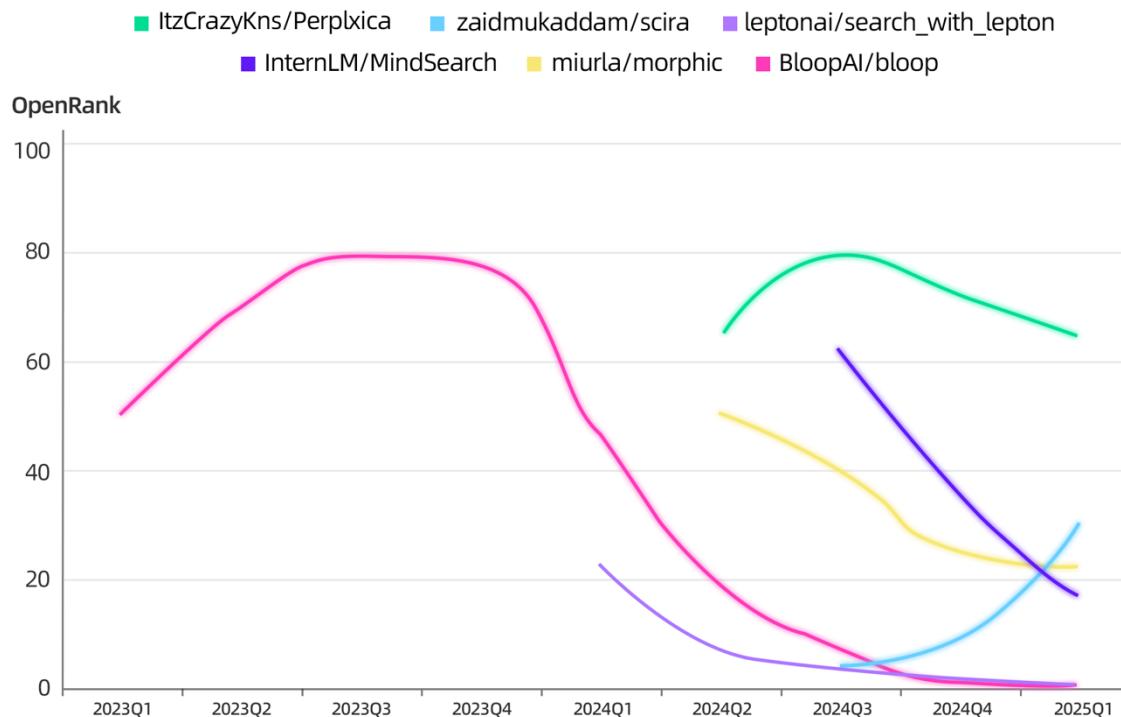
注：以上数据截止 2025 年 5 月 20 日



# 模型能力对应用场景的冲击与重塑

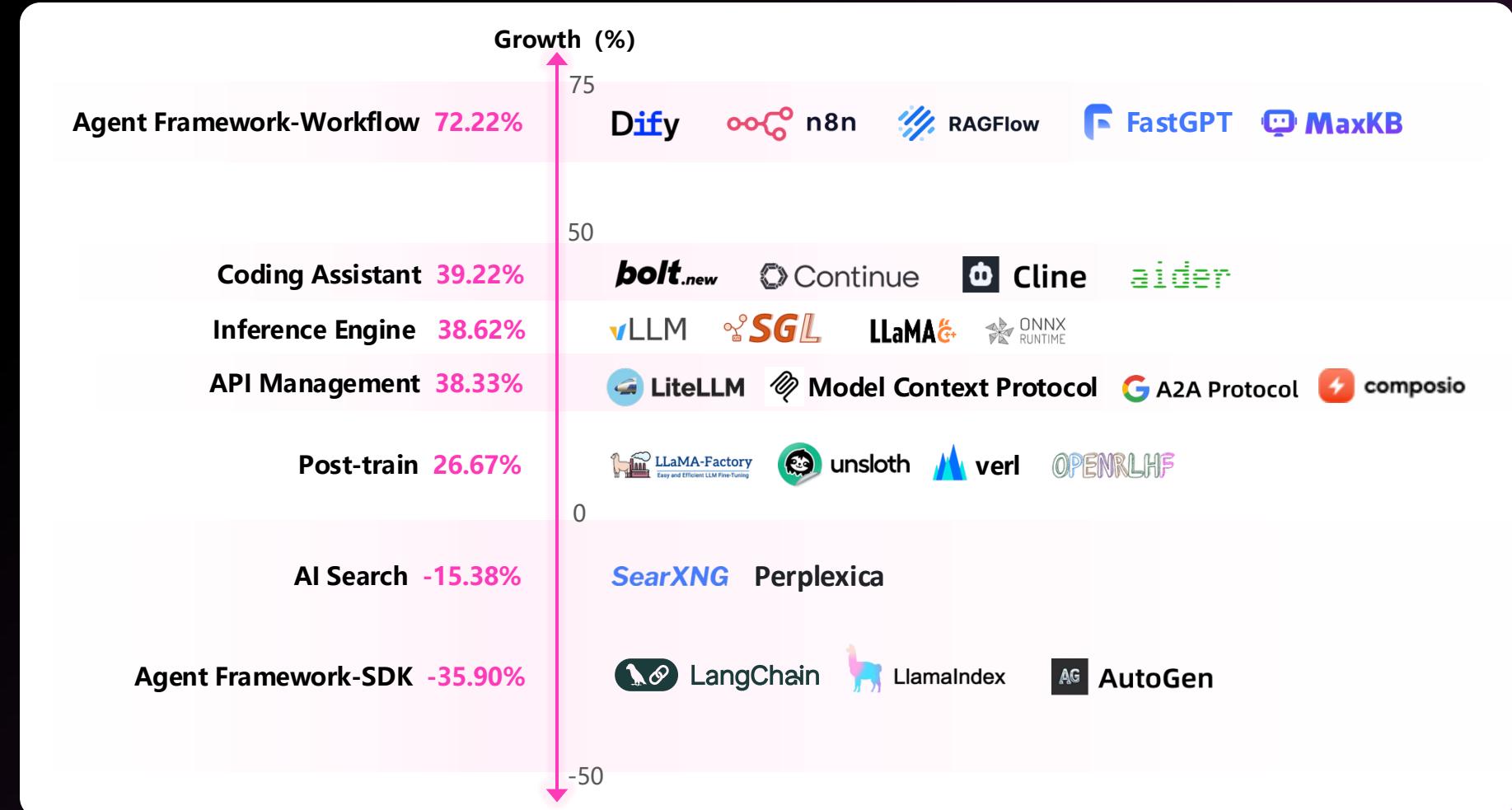
随着模型能力的不断提升，应用开发生态也在随之发生变革，有冲击与挤压，但也带来了更多新的想象空间。

- **AI Search 开源项目的式微：**AI 搜索是最早落地的应用场景，以 Perplexity 为代表的产品一度形成对谷歌搜索的实质性挑战，Morphic.sh、Scira 等开源项目也试图通过本地化部署和 API 自由配置打破闭源产品的垄断；从发展趋势上来看，这类项目的存活率并不高。
- **AI Coding 开源项目的火热：**与之相对的，模型能力的提升正在引发软件开发范式的变革。除了商业化产品 Cursor、Windsurf 等验证了市场热情外，以 Continue、Cline 为代表的 IDE 插件形态的项目们也是主流的开源选择，这两个项目的社区参与者人数都已经超过 3000 人，且 OpenRank 曲线持续攀升。



# 生态位之间的动态博弈

- 在增长的两端分别是以 Dify 为代表的低代码工作流式的应用框架和以 LangChain 为代表的 SDK 式应用框架。不同开发模式表现出的“冰火两重天”态势，也是该生态极快迭代与极强生命力的体现。
- DeepSeek-R1 推理模型带来的“Aha Moment”证明了强化学习这一后训练路径的有效性，以 Verl、OpenRLHF 为代表的强化学习框架在今年显著增长。



声明：所有 logo 均为引用，版权归其所在组织所有

注：分别取 2024 年和 2025 年 1 月至 4 月技术领域下所有项目的 OpenRank 均值，计算同比增长率

# 2025 大模型开源开发生态七大技术趋势



## 智能体应用

1. 2025 年, Agent 框架热潮褪去
2. 标准协议层: 兵家必争的战略要塞
3. 口嫌体直的 Vibe Coding 新范式



## 模型基础设施

4. 向量索引与存储的技术边界流动
5. 多模态数据治理在大模型时代的技术演化
6. 模型服务部署与推理的混战仍在持续
7. 以 PyTorch 为核心的训练生态

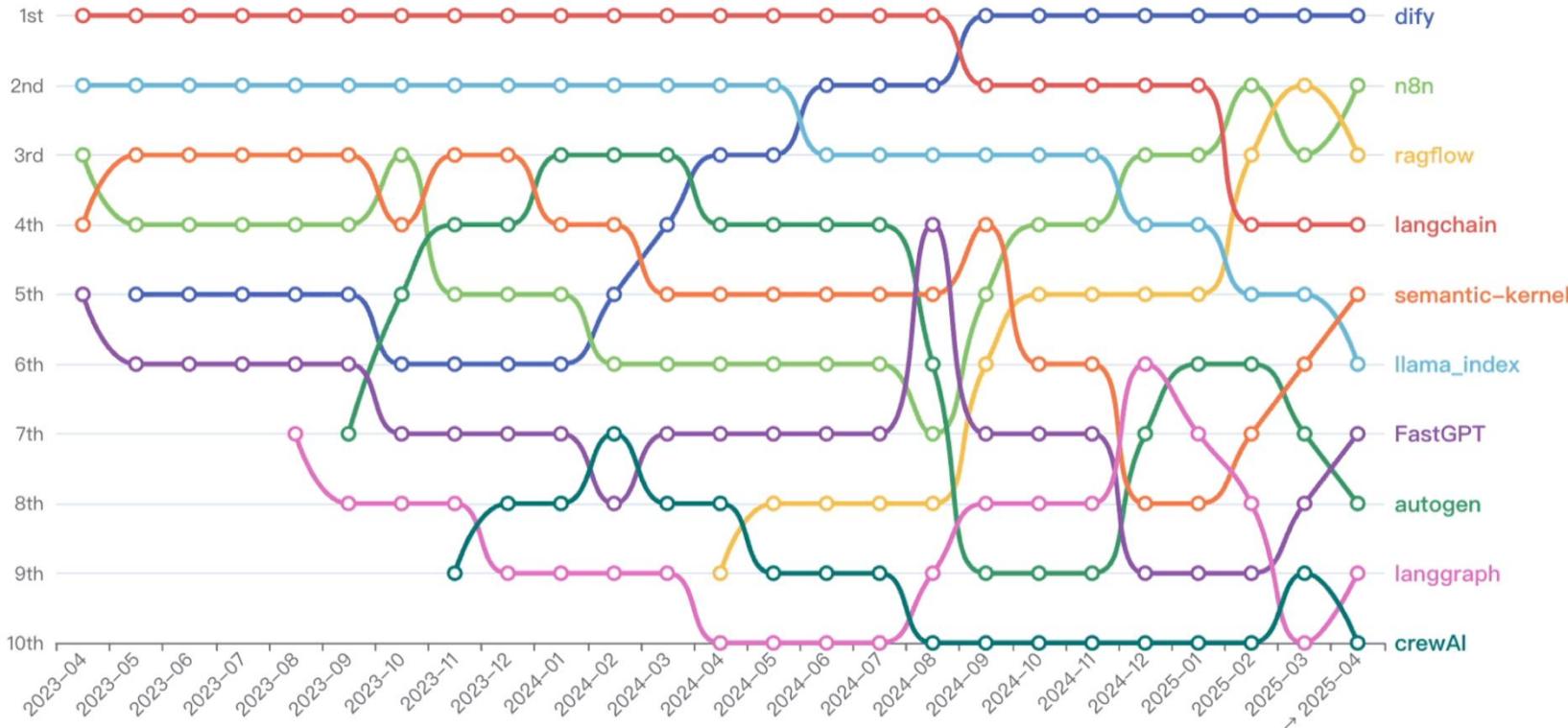
2025, Agent 框架热潮褪去

# Agent Framework 领域的热门项目们

Agent Framework 项目 OpenRank 排名			
		Star: ★	OpenRank: 🔥
01	 <b>Dify</b>	95,414	473
02	 <b>n8n</b>	88,153	213
03	 <b>RAGFlow</b>	51,009	208
04	 <b>LangChain</b>	106,738	159
05	 <b>Llamaindex</b>	41,353	120
06	 <b>AutoGen</b>	43,956	103
07	 <b>Semantic Kernel</b>	24,267	103
08	 <b>FastGPT</b>	23,852	89
09	 <b>LangGraph</b>	12,166	87
10	 <b>crewai</b>	30,840	76
11	 <b>MaxKB</b>	16,418	74
12	 <b>Agent Development Kit</b>	8,076	73
13	 <b>OpenAI Agents SDK</b>	9,740	69
14	 <b>CAMEL-AI</b>	12,233	64
15	 <b>Flowise</b>	37,827	53
16	 <b>AUTOgpt</b>	174,960	48
17	 <b>GraphRAG</b>	24,913	43
18	 <b>DB-GPT</b>	16,331	39
19	 <b>mem0</b>	28,389	34
20	 <b>haystack by deepset</b>	20,519	33
21	 <b>Letta</b>	16,279	27
22	 <b>MetaGPT</b>	55,207	17
23	 <b>LangChain-Chatchat</b>	34,878	16

# 2025 年, Agent 框架热潮褪去

Agent 框架 OpenRank Top10 排名变化



- 2023-2024 年间, 以 LangChain 为代表的“全能型”框架凭借其开创性的任务编排能力和丰富的工具集成一度主导市场, 无数“Agent Frameworks”在不同的技术概念和浪潮下迅速崛起;
- 到 24 年下半年, 格局初步形成, 已经少有看到新的开发框架出现, 热潮褪去之后, 我们看到早期主导市场的 LangChain 因为陡峭的学习曲线和复杂的调试成本在明显的走向下坡;
- 进入 2025 年, 格局呈现出分化的趋势: Dify、n8n、RAGFlow 等平台通过低代码工作流和企业级功能开始主导市场; 而 LangChain 这类开发框架逐渐式微。

标准协议层：**兵家必争的战略要塞**

# 标准协议层：兵家必争的战略要塞

## 2022 年, ChatGPT 兴起

早期的大模型应用开发者借助在 Prompt 中附上函数说明，来试图进行工具调用

## 2023 年, Function Calling 上线

OpenAI 在发布 GPT4-0613 模型时同时推出了 Function Calling 功能

## 2024 年 11 月, Anthropic 开源 MCP 协议

MCP (Model Context Protocol) 协议解决跨不同模型之间工具调用的标准化和上下文处理问题

## 2025 年 4 月, Google 开源 A2A 协议

A2A (Agent2Agent) 协议促进智能体之间的通信和协作

## 2025 年 5 月, CopilotKit 开源 AG-UI 协议

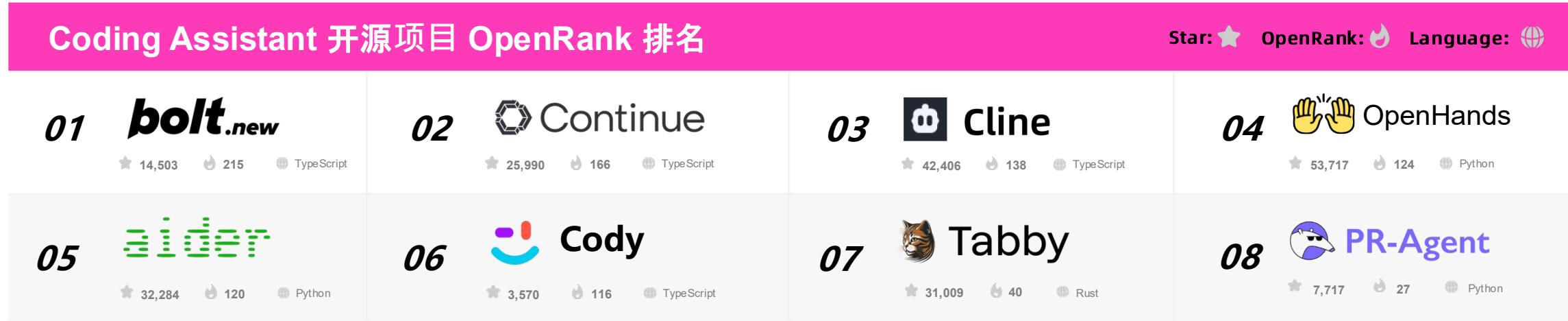
AG-UI (Agent-User Interaction Protocol) 协议是一种智能体和用户之间的交互协议



随着大模型服务的加速演进，**标准协议层也会成为头部玩家的战略要塞**，在可见的未来一到两年，标准协议层可能迎来密集的生态卡位战，而那些深度融合**技术前瞻性与开发者体验**的协议框架将逐渐占据生态话语权，并引导大模型技术领域创新的规模化涌现。

口嫌体直的 Vibe Coding 新范式

# 口嫌体直的 Vibe Coding



注: 排名取 2025 OpenRank 月均值, 数据截止 2025 年 4 月 30 日

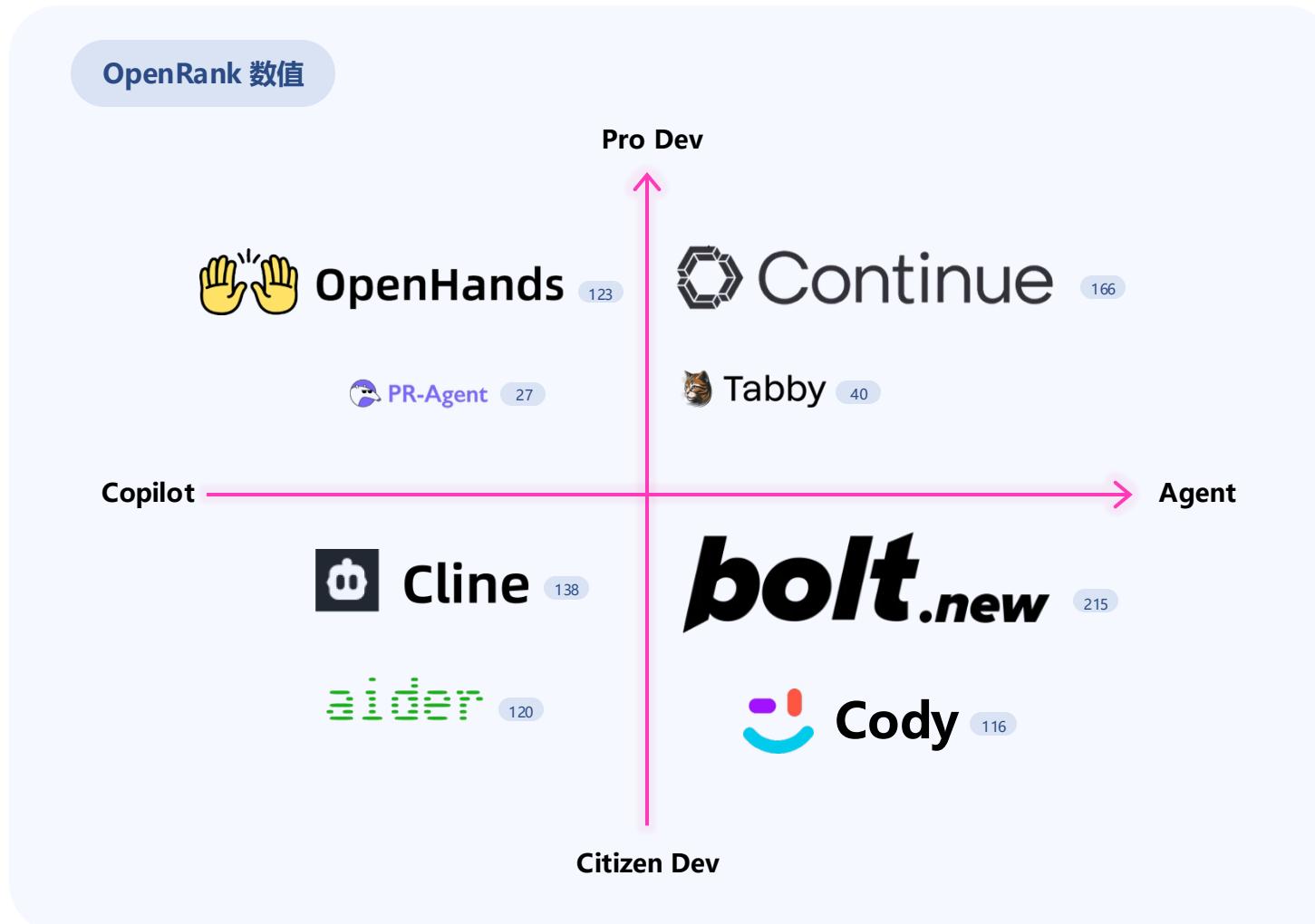
程序员会率先被 AI 取代吗？再往前几个月，这样的问句频繁出现。当 GitHub Copilot 帮助做代码自动补全时，开发者们产生了存在主义危机；当 AI IDE 们直接上手创建完整的工程项目时，真香定律上线，氛围编程（Vibe Coding）成为新型的开发模式，一大批开源的 AI 辅助编程项目们也打得火热。

## ■ 大厂在代码辅助编程赛道快速下场，但几乎以闭源为主

GitHub Copilot、Amazon Q developer、CodeArts Snap（华为）、通义灵码（阿里）、Trae（字节）、CodeFuse（蚂蚁）

## ■ 反而是一些创业公司或三五人的小团队能够快速产出一个该领域的开源项目，迅速出圈

# 口嫌体直的 Vibe Coding

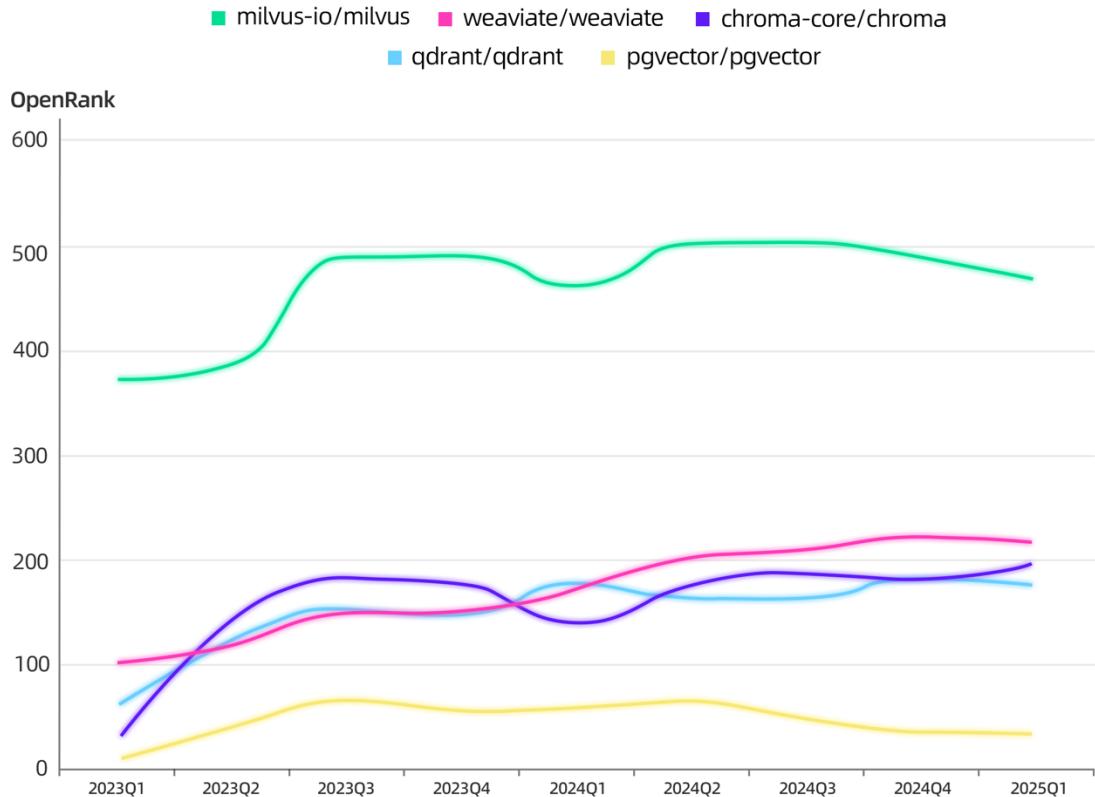
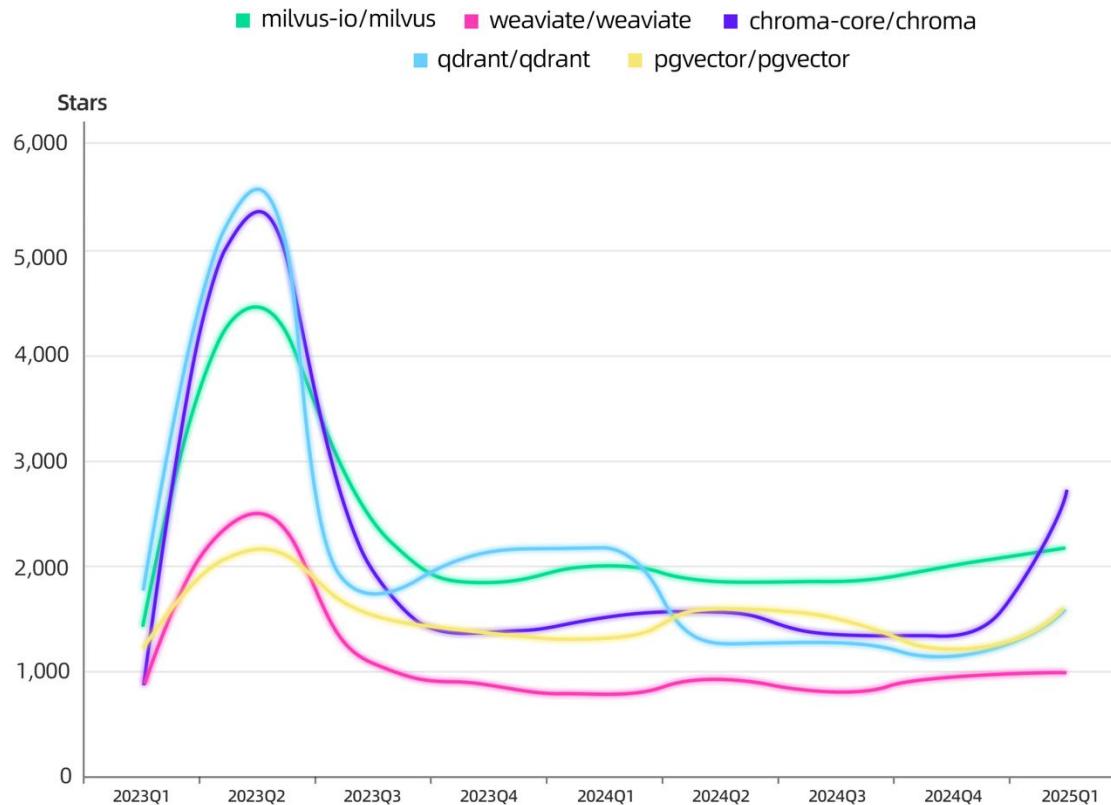


AI Coding 正在尝试从一次性代码生成走向真实的软件工程场景下的开发，但依然面临很大挑战

预计未来 24 个月内，随着代码验证技术（如形式化方法与符号执行的结合）、多模态训练数据（代码+文档+运行时日志）的成熟，以及开发者反馈闭环的优化，AI 开发助手将会承担更多常规开发任务，但仍需人类开发者在关键决策点进行监督。

# 向量索引与存储的**技术**边界流动

# 向量索引与存储的技术边界流动



- 向量数据库的发展“起于漫天的富贵，回归理性的沉淀”。2023年2月前后，以Qdrant, Chroma为代表的项目收到了一波极其夸张的关注，分别收获了超过5000的Star数，但这种高关注并未形成持续的趋势。在2024-2025年的整体发展期，各个项目新获得的关注整体趋于稳定，没有出现很大的差异；
- 在OpenRank趋势上，开源较早并且由LF AI & Data基金会中立托管的Milvus一直在保持着稳定的领先，而整个技术领域似乎像是平行的跑道向前发展。

# 向量索引与存储的技术边界流动

这种状态的背后的原因有几个关键的可能性：

- **开源产品并非唯一的市场选择：**同期有纯商业版的强力竞争者如 Pinecone, KDB.AI 出现，且 Pinecone 的产品力和市场拓展相当不错；
- **传统大数据系统的向量化升级：**随着时间的推移，传统数据库开始推出向量化的插件或向量搜索引擎（如图中的 pgvector），这当中包含了大量很受欢迎的 DB 选型，如 PostgreSQL, MongoDB Atlas, Elasticsearch 等，对于纯粹的向量数据库带来了一些冲击；
- **OpenCore 模式下，商业公司不那么关心 Core 的数据活跃：**正如传统数据库有大量的项目是通过「开源核心」的方式来做商业模式，向量数据库也是类似。而这种模式下，有一个可用完整的开源核心非常重要，但核心的整体活跃度并不是背后的商业公司最关注的事情，基于核心版本形成的有竞争力的生态才是。

这带来了另一个讨论：

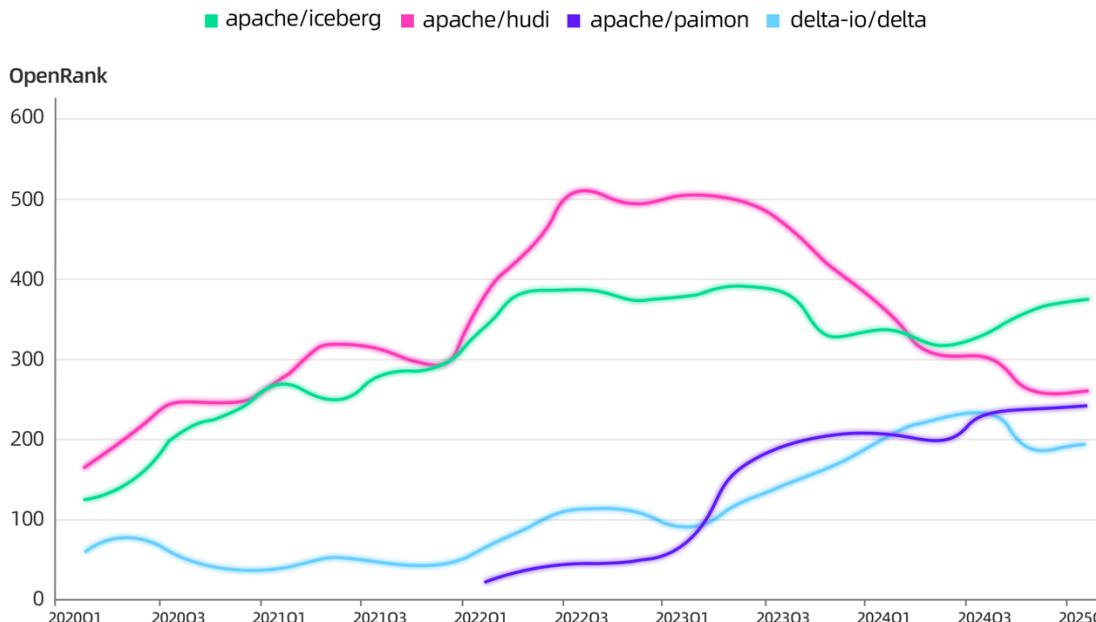
**向量数据库是不是一个“冗余技术”？是否用传统数据库结合 VSE 就可以满足模型应用的需求场景？**

- 答案是没有。社区数据显示，pgvector 的趋势处于不升反降的状态

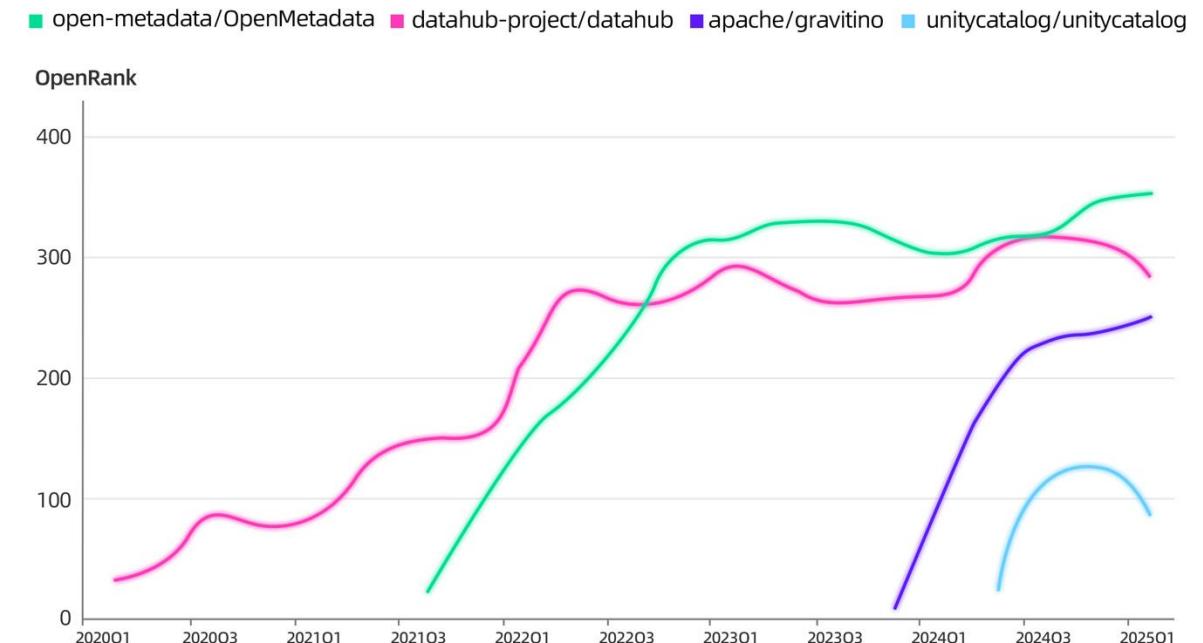
# 多模态数据治理在大模型时代的技术演化

# 多模态数据治理在大模型时代的技术演化

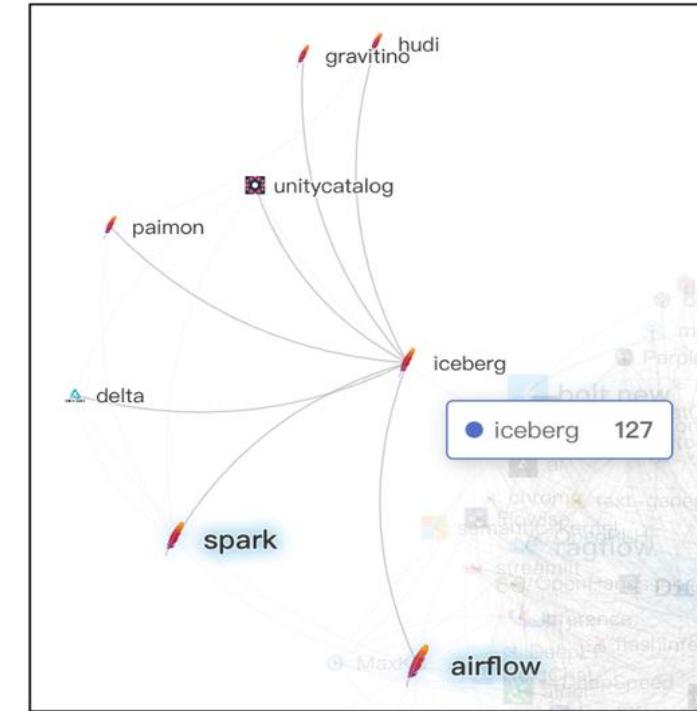
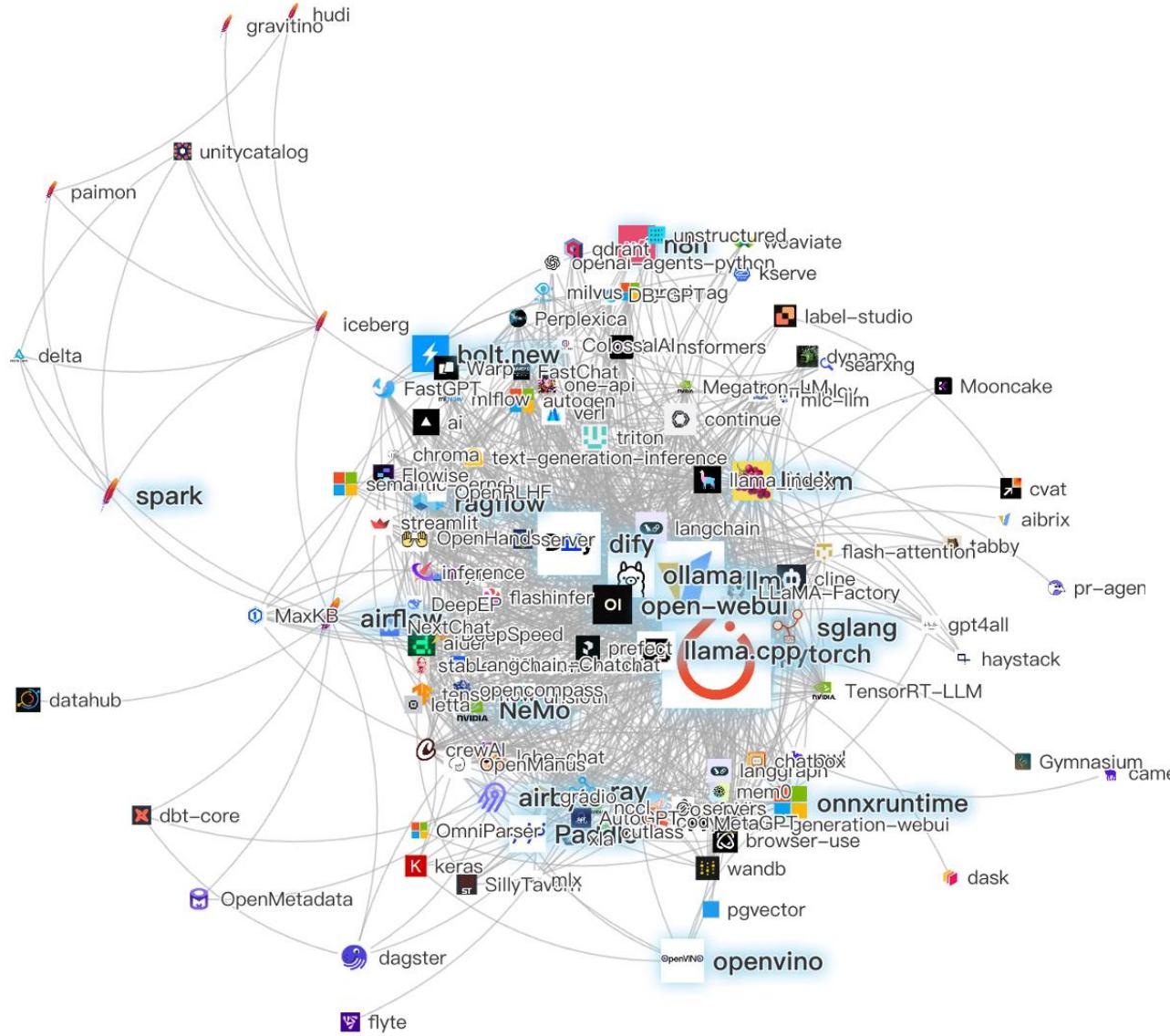
## 数据湖表格式方面



## 元数据与数据目录方面



# 多模态数据治理在大模型时代的技术演化

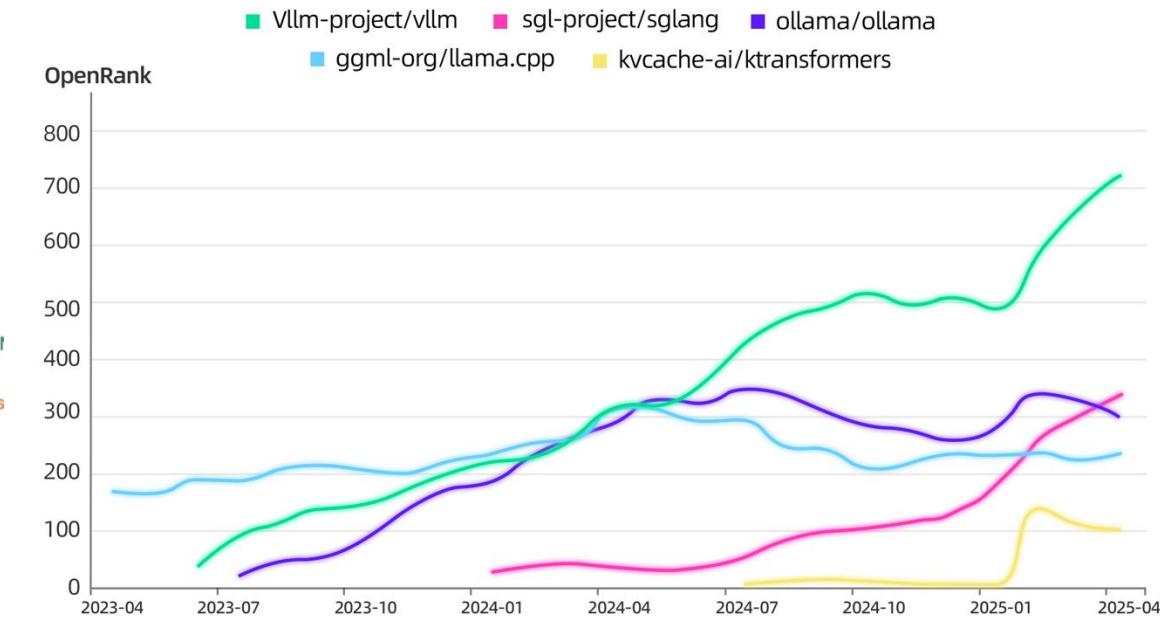
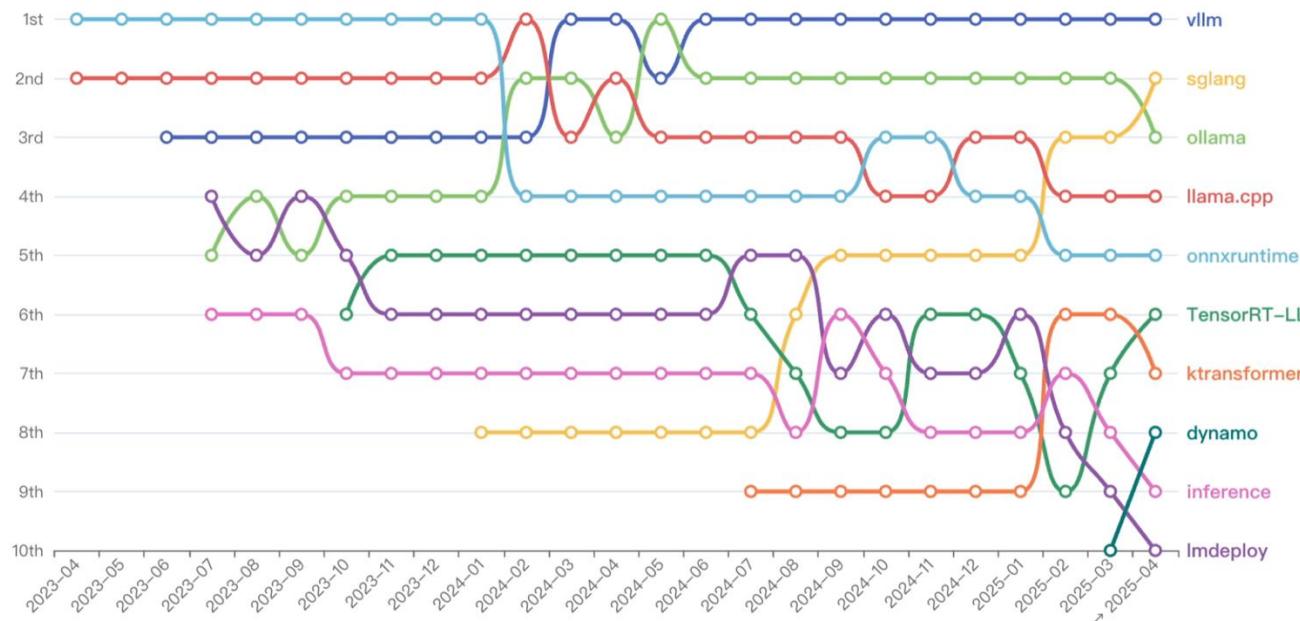


- 从生态协作关联来看，大数据系统领域的项目们处在开发生态相对边缘的位置，推测数据和 AI 开发生态的融合还在路上；
- 在未来，随着基座模型进一步融入数据基础设施，各项目之间的联动也许会更加紧密，无论是为机器学习任务提供高质量数据，还是利用模型反向的支持数智化的数据治理。

模型服务部署与推理的混战仍在持续

# 模型服务部署与推理的混战仍在持续

- vLLM 和 SGLang 是当下大模型领域最著名、热度最高的两个推理引擎，也都是 DeepSeek 推荐使用的推理引擎
- 2024 年第四季度，vLLM 的增长一度趋于停滞，而同期 SGLang 正在快速迭代，其 Q4 OpenRank 平均增速达 12%。进入 2025 年后，vLLM 发布了 v1 重大版本，完成核心架构升级，重新进入增长通道。推理引擎生态似乎由此开启了新一轮“AI 军备竞赛”：2025 年第一季度，vLLM 的 OpenRank 平均增速为 17%，SGLang 则高达 31%



# vLLM 与 SGLang 社区数据对比

	vLLM	SGLang
项目影响力 (OpenRank)	25 年均值 : 615 25 年增速 : 17%	25 年均值 : 269 25 年增速 : 31%
发起组织	SkyLab, UC Berkeley	LMSYS, 一个由 UC Berkeley, Stanford, CMU 多家大学共同发起的开放式研究性组织
重要开发者	<a href="#">DarkLight1337</a> , HKUST, 香港 <a href="#">youkaichao</a> , 清华 · 北京 <a href="#">WoosukKwon</a> , UCB, 加州	<a href="#">merrymercy</a> , xAI, 加州 <a href="#">zhyncs</a> , Baseten, 加州 <a href="#">zhaochen yang20</a> , UCLA, 加州
开发者分布	美国 (33.2%) , 中国 (30.8)	中国 (49.2%) , 美国 (27.5%)
社区规模	贡献者 : 449 参与者 : 12332	贡献者 : 1134 参与者 : 2232

注: 以上数据截止 2025 年 5 月 20 日

继 Ray 和 Spark 之后, UC Berkeley 又一次展示了其强大的顶尖开源技术的孵化能力:

- vLLM 诞生于 UCB SkyLab 实验室, 和 Spark、Ray 一脉相承;
- SGLang 诞生于 [LMSYS](#), 是 UCB 和多家大学共同发起的研究性组织, 该组织同时还创立了火爆的大模型对抗式评测平台 Chatbot Arena。

# 并不只有 vLLM 与 SGLang

Model Serving 项目 OpenRank 排名					Star: ★	OpenRank: 🔥	Language: 🌎
01  vLLM	02  Ollama	03  OpenVINO	04  SGLang	05  LLaMA	★ 46,409	🔥 615	Python
06  ONNX RUNTIME	07  Tensor RT-LLM	08  KTransformers	09  LMDeploy	10  Xorbits	★ 139,289	🔥 311	Go
11  NVIDIA Dynamo	12  Text Generation Inference	13  Triton Inference Server	14  gpt4all	15  KServe	★ 8,220	🔥 280	C++
16  Mooncake	17  AI Brix	18  MACHINE LEARNING COMPILATION	19  BentoML		★ 13,902	🔥 269	Python
16,451	10,401	13,859	6,274	7,711	201	96	C++
3,923	10,076	9,154	73,241	7,126	63	41	Rust
3,170	3,511	20,532	7,667	4,126	20	18	Jupyter Notebook

注: 排名取 2025 OpenRank 月均值, 数据截止 2025 年 4 月 30 日

但模型服务这条赛道上，并不只有 vLLM 与 SGLang :

- **Ollama 与 llama.cpp** : 端侧推理和本地部署的轻量级优选。使用 llama.cpp 进行模型训练、量化和性能调优，再通过 Ollama 进行快速部署和服务化管理，在个人开发者中是一种常见实践。这两个项目不仅仅是在模型服务赛道最近一个月的排名上位列第 3 和第 4，从整个生态来看，也占据不可忽视的头部位置。
- **KTransformers** : 超大规模参数场景下的异军突起。2025 年 2 月，清华大学 KVCache.AI 团队推出的 KTransformers 宣布成功在 24 GB 显存 + 382 GB 内存的 PC 上实现本地运行 DeepSeek-R1、V3 的 671B 满血版，速度提高 3~28 倍，随即迎来社区爆发式增长—当月 OpenRank 飙升 34 倍，吸引了 736 名开发者参与协作与讨论，Star 数突破 1 万。

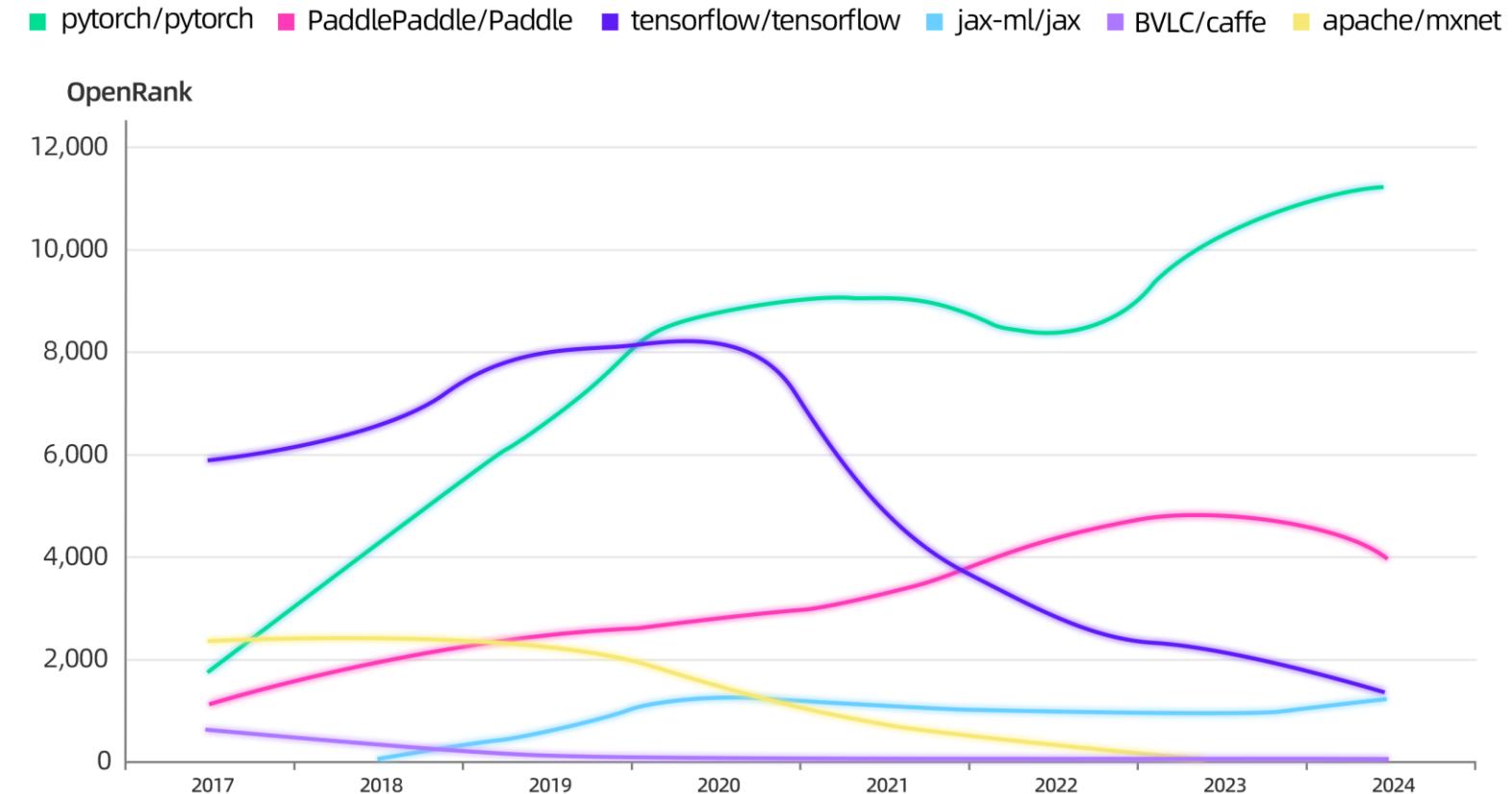
以 PyTorch 为核心的训练生态

# 以 PyTorch 为核心的训练生态



# 以 PyTorch 为核心的训练生态

PyTorch 是如今大模型开发生态的绝对顶流和生态核心，也是模型训练领域当之无愧的事实标准。凭借模块化和轻量化的开放设计，在2020 年正式超过了 TensorFlow，成为大模型时代深度学习的基础设施，而 TensorFlow、MXNet、Caffe 等框架已经成为了上一个时代的过去式。



# 发布前一周，还发生了什么？

## Microsoft Build 2025 大会重磅发布

- Windows 操作系统原生支持模型上下文协议 (MCP)
- 微软开源 NLWeb：“智能体 Web 时代的 HTML”
- VSCode 开源 GitHub Copilot Chat 插件，推动 AI 编程发展

## Google I/O 2025 大会重磅发布

- Gemini 系列升级，正在成为一个“AI 操作系统”，将被整合到安卓各种设备和 Chrome 浏览器里
- 谷歌“开放”AI 模型家族 Gemma 继续壮大：Gemma 3n, MedGemma 和 SignGemma（即将推出）

## Anthropic Code with Claude 2025 大会，Claude 4.0 发布

- Claude 4.0 发布，7 小时不停歇编程，AI 应用的产品和开源项目们又开始了新一轮黑客松

# 附录

- **Landscape** 发布地址：

<https://antoss-landscape.my.canva.site/>

- **Landscape** 完整项目列表和数据：

<https://docs.google.com/spreadsheets/d/1av9kitgnRGtsmDp6AbW96m2cCR4jXZFQmUVG2di8Bjw/edit?gid=0#gid=0>

- **OpenRank** 指标介绍：

[https://open-digger.cn/docs/user\\_docs/metrics/global\\_openrank](https://open-digger.cn/docs/user_docs/metrics/global_openrank)

虽然我们非常努力想从中挖掘更多信息，但我们也完全明白，社区的数据既不全面也不完全准确，而且也不一定能反映出很多最新最优秀的技术变化，我们只希望这个报告能给大家一些有益的参考，有什么错漏之处和其他值得补充的观点，也欢迎大家反馈给我们。

# Thanks!

关注蚂蚁开源公众号推送，获取完整版报告与解读

