QCon
全球软件开发大会

从社区数据看大模型开发生态全景与趋势

演讲人：王旭，夏小雅

蚂蚁集团 / 开源技术委员会副主席，开源技术增长经理

蚂蚁开源
ANT OPEN SOURCE

关于我们

王旭

写开源Bug，做开源项目，搞开源创业，混开源社区，管公司开源，参与开源基金会。

分析开源数据，发开源论文，搞开源学术，写开源代码，管开源社区，做公司开源策略。

夏小雅

... 大人，时代变了，大模型来了

目 录

01 案例：热点到数据

02 全景：社区关系网

03 热点：项目与趋势

04 架构与全景图

InfoQ 极客传媒

AiCon 北京站
全球人工智能开发与应用大会

AI Agent

RAG

多模态

关于 AI 的高频问题 都能在这里找到答案

AI 智驾

从大模型变革之路到高效“炼丹”指南

成本优化实践

扫码领取你的智囊团

AI Native 产品创新
与技术落地

咨询购票

查看详情

01

案例研究

从热点新闻到社区数据

社区的行为数据只反映一部分的社区关系，既不全面，也不充分，并不被视为是对项目本身「好坏」的公平评价，我们只是以此来提供一些参考信息。

我们怎么观察 AI 开发生态

ANT OPEN SOURCE

从GitHub上，我们可以有的行为信息

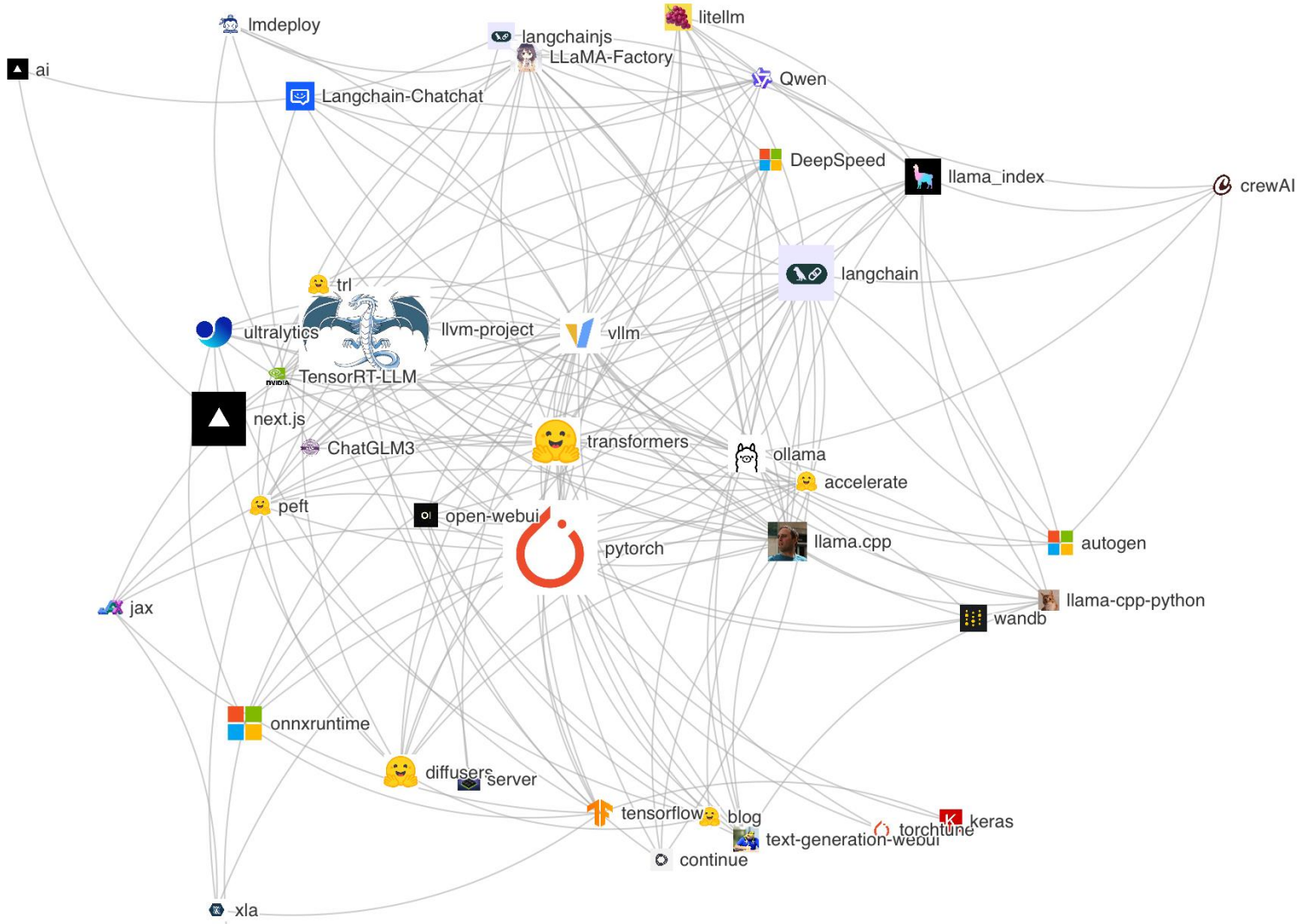
- 项目本身的活动数据。
  - 项目本身的 Commit / PR / Issue 数量、频率、分布情况等
- 项目之间的彼此关联信息——项目之间的互动——通过共同的参与人达到的互动。
  - 同时对两个项目的 Commit / PR / Issue / Comments 等。
  - 以及 Star/Fork （以下分析中没有使用这两个数据）

项目之间可能有的生态关系

- 依赖性
  - 有依赖性的项目，有可能会有互相的，尤其是对被依赖项目的贡献
- 竞争性/可替换性
  - 同一位置的项目，有可能会有上下游同时给他们贡献
- 上下游（合作）
  - 合作者有可能会有彼此的贡献

* 这里都是有可能，但不是必然的。

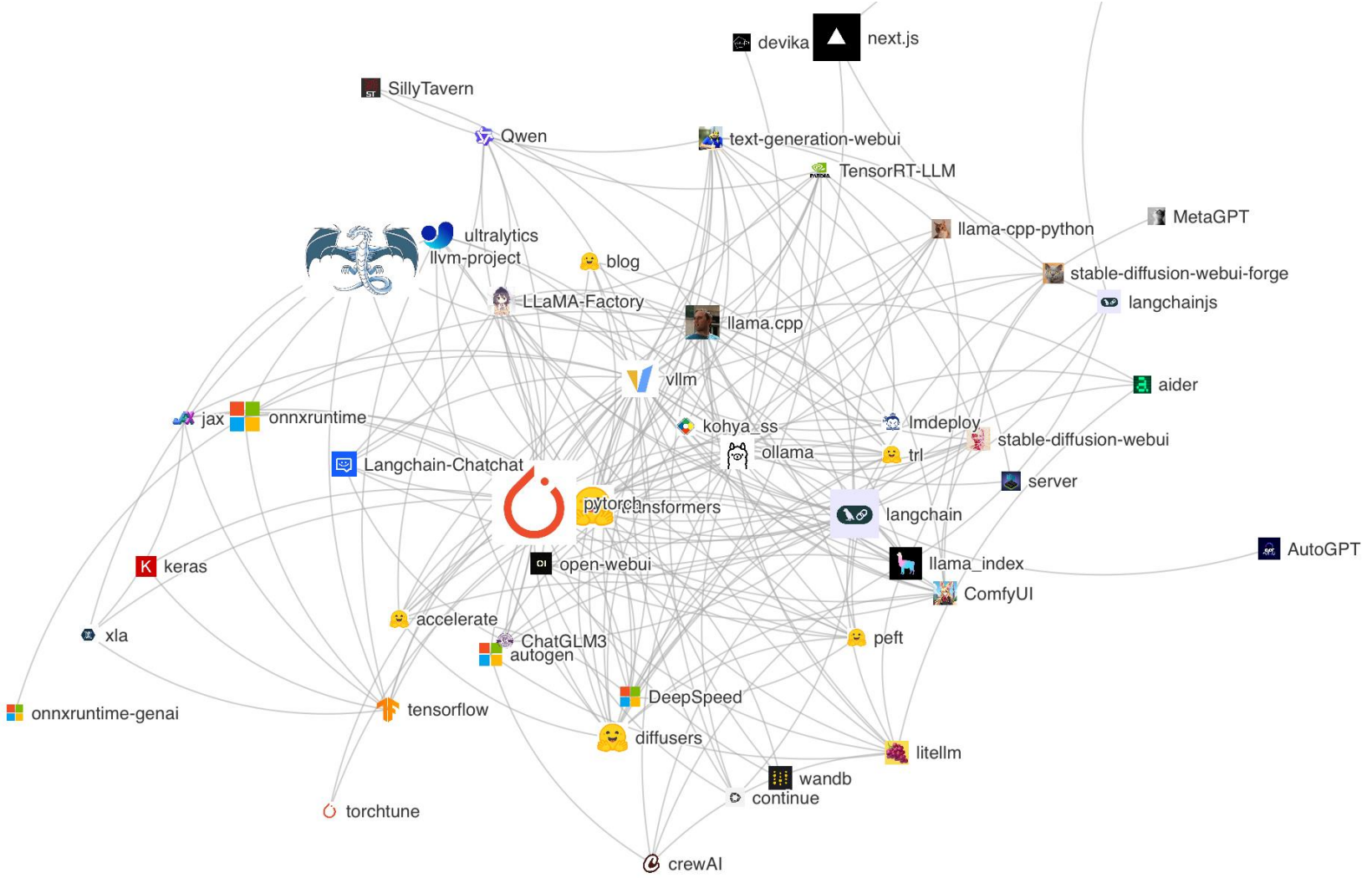
从 PyTorch, vLLM, LangChain 出发



https://xiaoya-yaya.github.io/ant-ospo-insights/

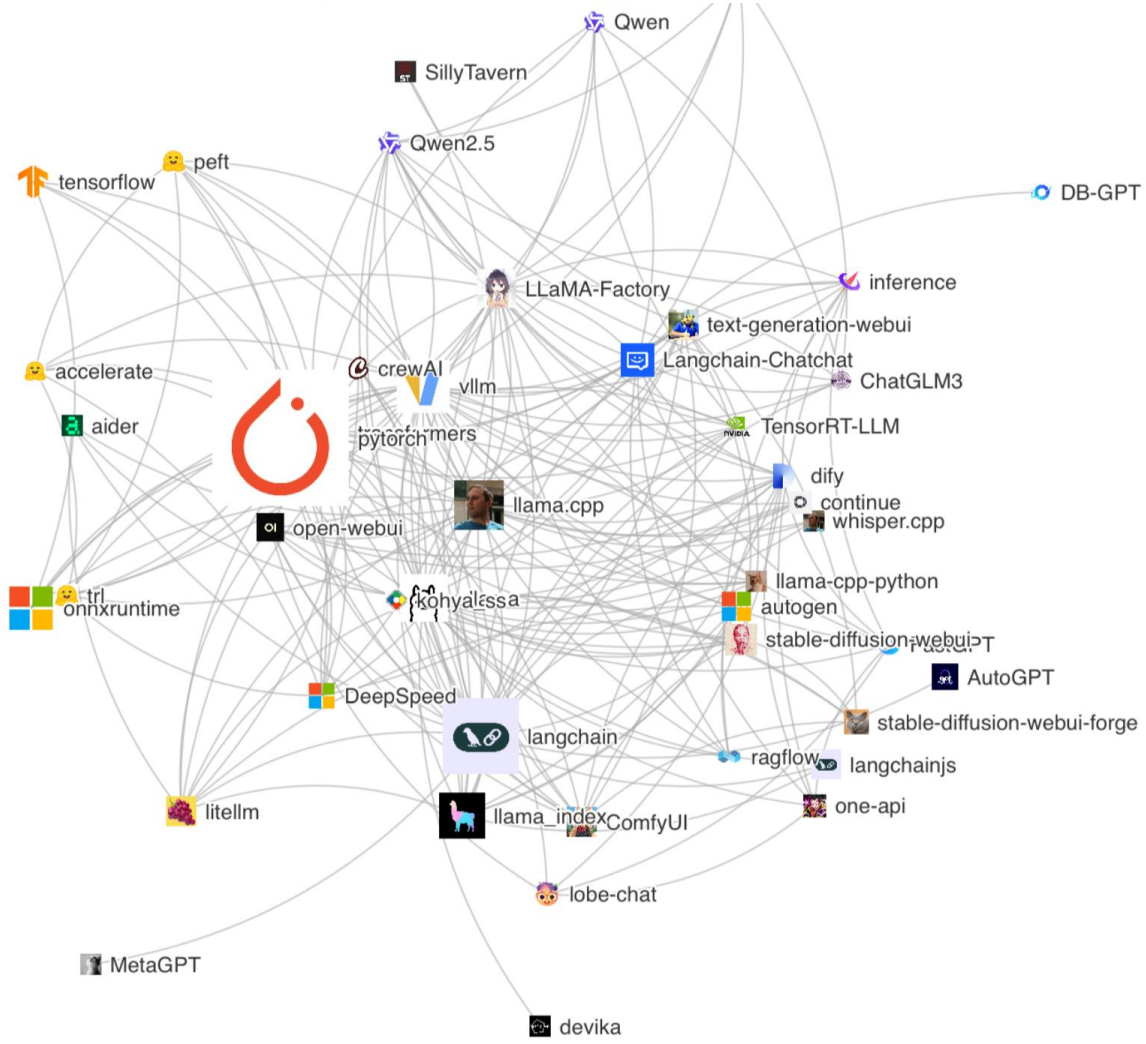
+AutoGPT

ANT OPEN SOURCE



https://xiaoya-yaya.github.io/ant-ospo-insights/

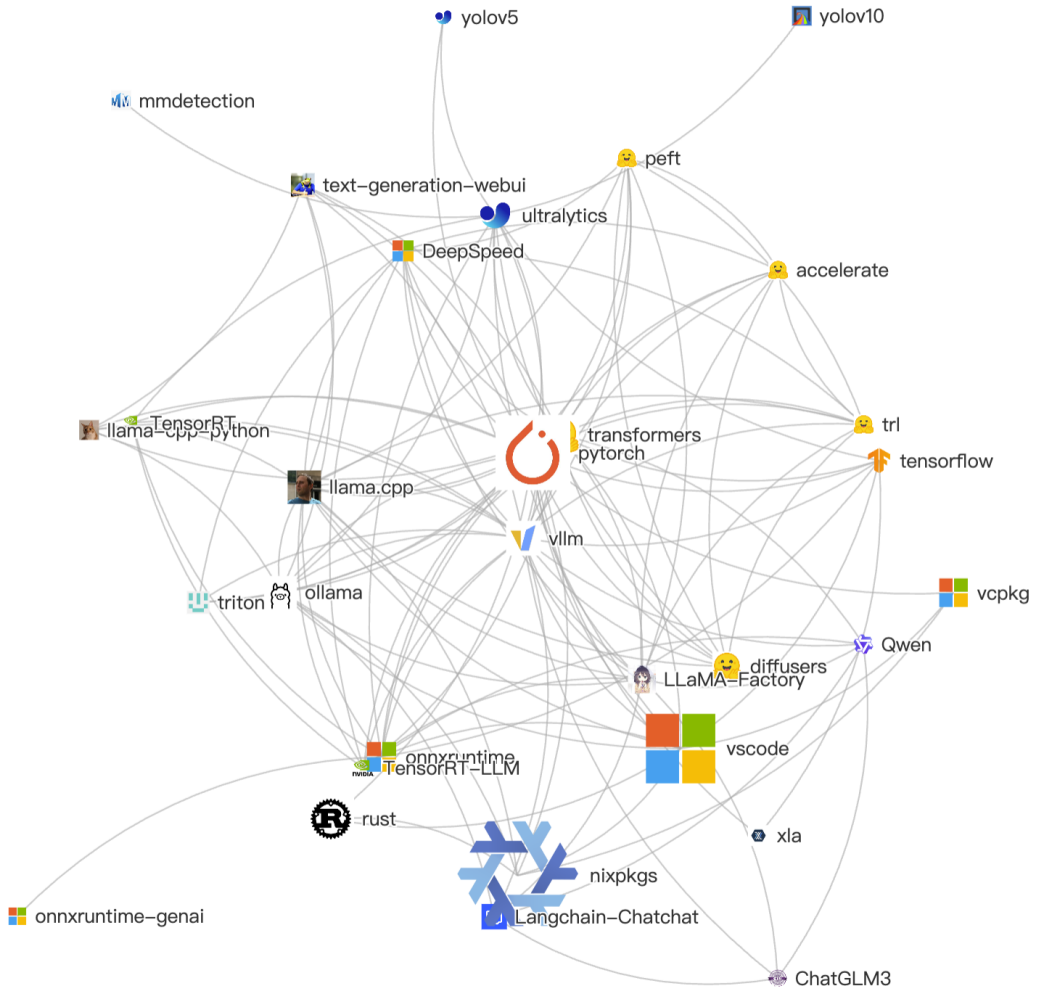
从 AutoGPT, llama.cpp, Chatchat 出发



https://xiaoya-yaya.github.io/ant-ospo-insights/

向下：从 DeepSpeed, Triton, ONNX 出发

ANT OPEN SOURCE



https://xiaoya-yaya.github.io/ant-ospo-insights/

插播一个有趣的发现

ANT OPEN SOURCE

从过去一年里项目名/仓库名发生变化的这些热门 AI 项目们：

- openai/triton -> triton-lang/triton
- OpenDevin/OpenDevin -> All-Hands-AI/OpenHands
- jmorganca/ollama -> ollama/ollama
- google/jax -> jax-ml/jax
- jerryjliu/llama_index -> run-llama/llama_index
- hiyouga/LLaMA-Efficient-Tuning -> hiyouga/LLaMA-Factory
- KillianLucas/open-interpreter -> OpenInterpreter/open-interpreter
- ollama-webui/ollama-webui -> open-webui/open-webui
- joaomdmoura/CrewAI -> crewAIInc/crewAI

多数为从个人项目发起，迅速火爆继而 transfer 到 GitHub ORG 下面和社区共同治理的，也有从如 Google 这样的企业生长出来演变成一个中立社区进行托管的，还有因为项目在演化过程中 roadmap 逐渐发生变化而修改项目名以重新定位公众认知的，如 LLaMA-Factory， OpenHands。

一些发现

ANT OPEN SOURCE

哪些项目会被引入进来

- 改变出发点会改变入选项目，但有些热点项目，不论从哪个点出发，都很容易被引入进来：
  - PyTorch, llama.cpp, huggingface/transformers
  - 可以认为关心 LLM 生态的人都会去这些项目有所表达；
  - 距离应用/算法开发比较近的项目热度会更高
- 开发者本身是会有扎堆现象的：
  - 比如中国开发者会更容易和中国开发者产生互动：
  - 比如 ChatChat 就更容易把 MetaGPT, DB-GPT Dify，FastGPT 这些项目给联系进来

哪些项目不会被引入进来

- 比如 Kubernetes，就完全没有被引入，虽然大家都会用它，我们推测：
  - API 已经比较稳定，即使是利用了接口也不需要去社区互动；
  - 同时快速发展的项目可能更容易有互动
  - 以及有很可能的一点——计算生态圈和AI生态圈的开发者之间可能是有相对强的隔离的
- 有些很多人用的项目，但是互动比较少：
  - 比如大部分 Nvidia 的项目都没有被引入进来，推测是单向使用，或者社区比较高冷。

02

生态全景

从社区关联到生态关系网

生态是本无所谓有，无所谓无的。正如这些公共接口；其实项目之间本没有公共接口，重复造的轮子多了，也便成了接口。

——这不是鲁迅说的，是王旭说的

全景图中用到的前置定义

ANT OPEN SOURCE

几个不同视角或线索

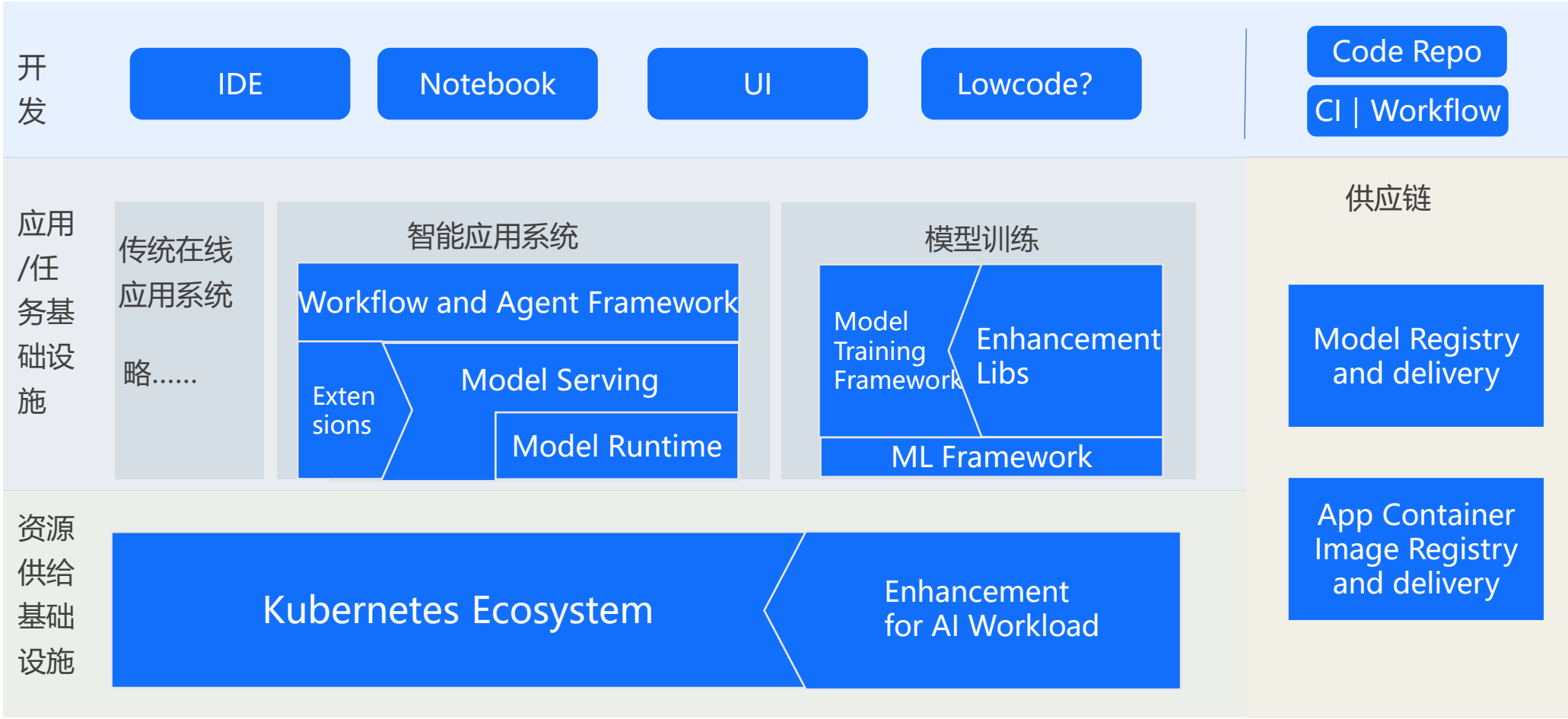
- 开发阶段
  - 从人的输入，到代码进入【代码仓库】；
- 供应链
  - 从代码或数据的存储，到镜像/模型注册中心，再到执行点
- 运维
  - 部署与升级：编排、调度、运行、更新等
- 运行阶段
  - 端到端工作流，上下游的配合

大模型生态中的变化点

- 围绕预训练模型
  - 分布式预训练本身是庞大的计算任务，而且是围绕 GPU 的，相关有很多项目出现；
  - 模型的庞大尺寸，催生了新项目，或者让相关的项目产生了变化；
  - 从数据库为中心的服务到模型为中心的服务；
- 开发和应用流程
  - 更注重自然语言等输入手段和工具本身的智能化——交互式、智能化的开发

AI生态中的生态位

ANT OPEN SOURCE



AI生态中的开源项目们

ANT OPEN SOURCE

一些可能的架构断言

ANT OPEN SOURCE

模型可能成为架构中的一等公民

- 模型成为“新的数据库”；
- 模型运行时可能成为和应用运行时并列的一等公民；
- 大模型的供应链上，大模型的尺寸，训练工作流和应用镜像相似又有不同；

开发工具可能会有更多变化

- 开发者工具，尤其是IDE，一直以来是最活跃的地方，也是 AI 可以产生影响很大的地方——有灵魂的工具；
- CI、测试领域可能还会有更多变化

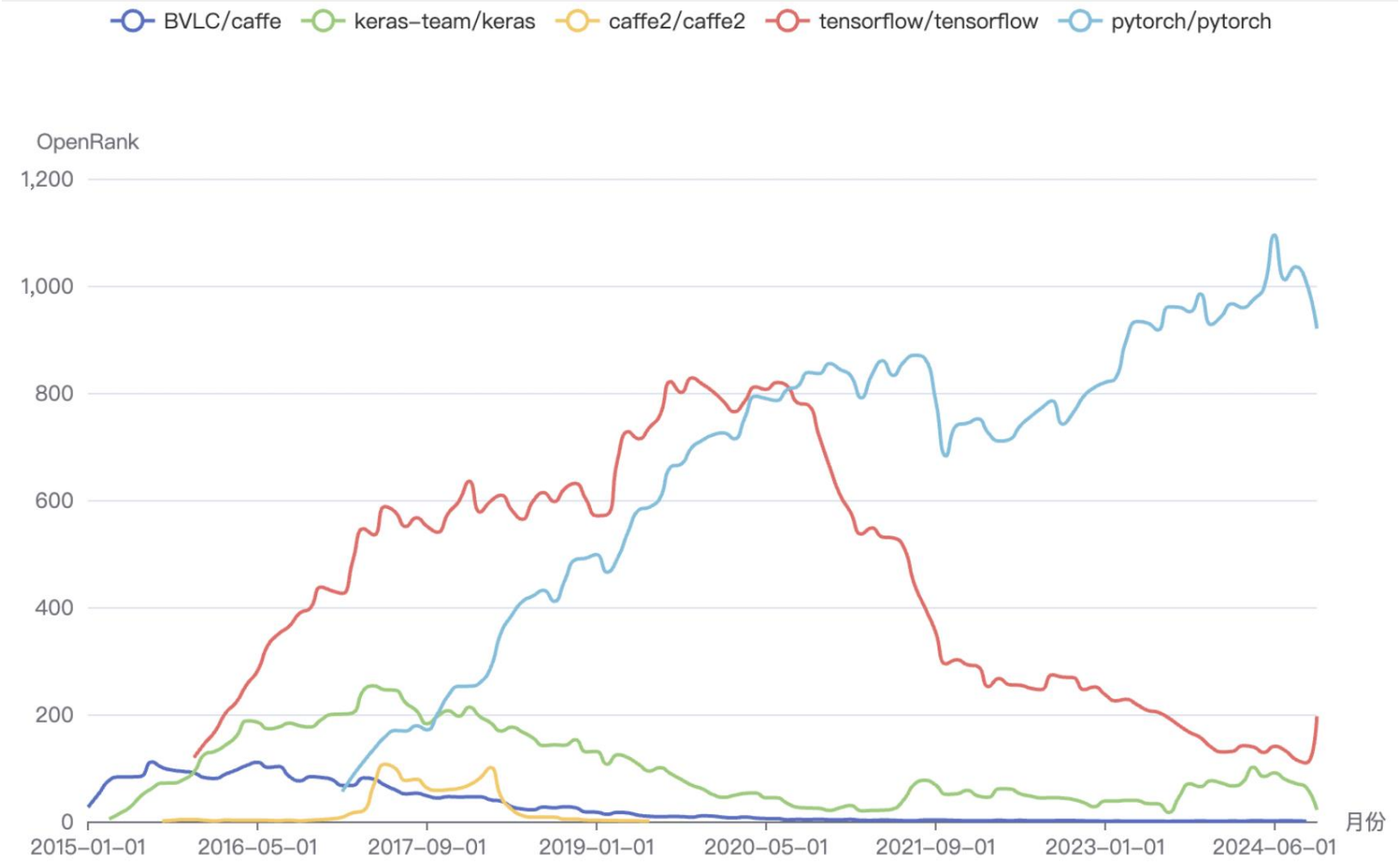
03

热点洞察

趋势上的项目与项目的趋势

传统机器学习框架的十年发展趋势

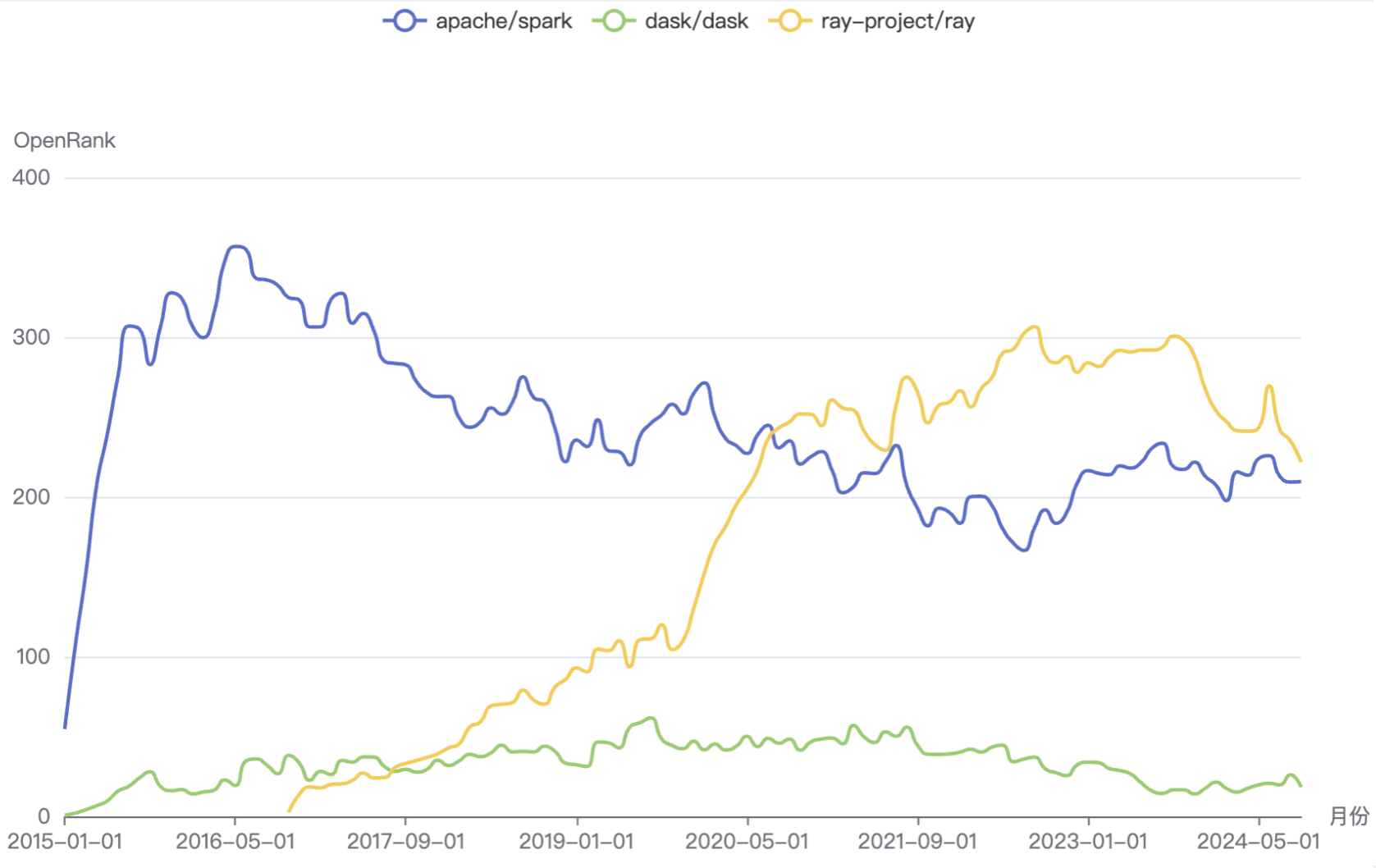
ANT OPEN SOURCE



https://xiaoya-yaya.github.io/ant-ospo-insights/

大规模分布式计算引擎

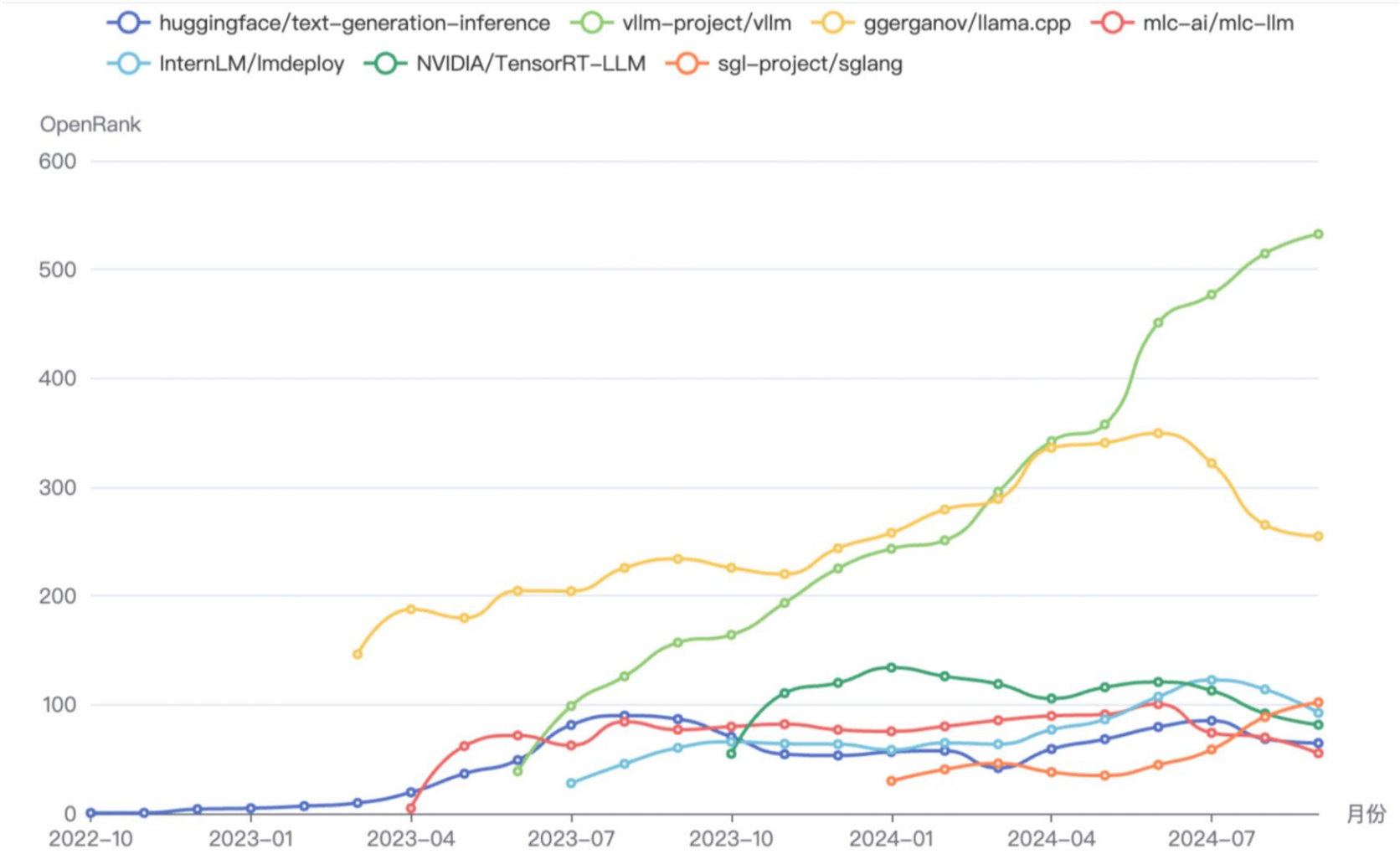
ANT OPEN SOURCE



https://xiaoya-yaya.github.io/ant-ospo-insights/

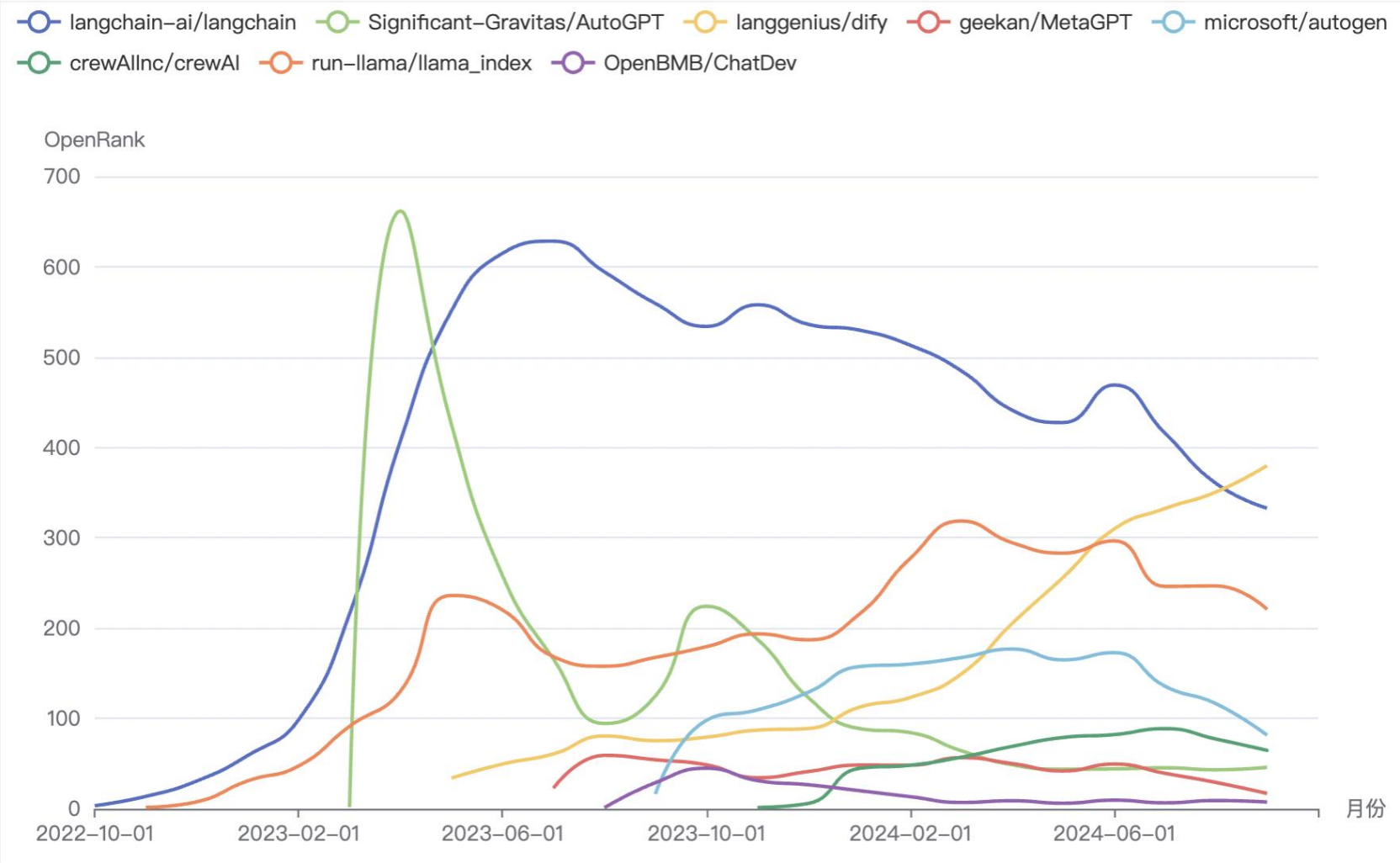
推理部署服务引擎

ANT OPEN SOURCE



https://xiaoya-yaya.github.io/ant-ospo-insights/

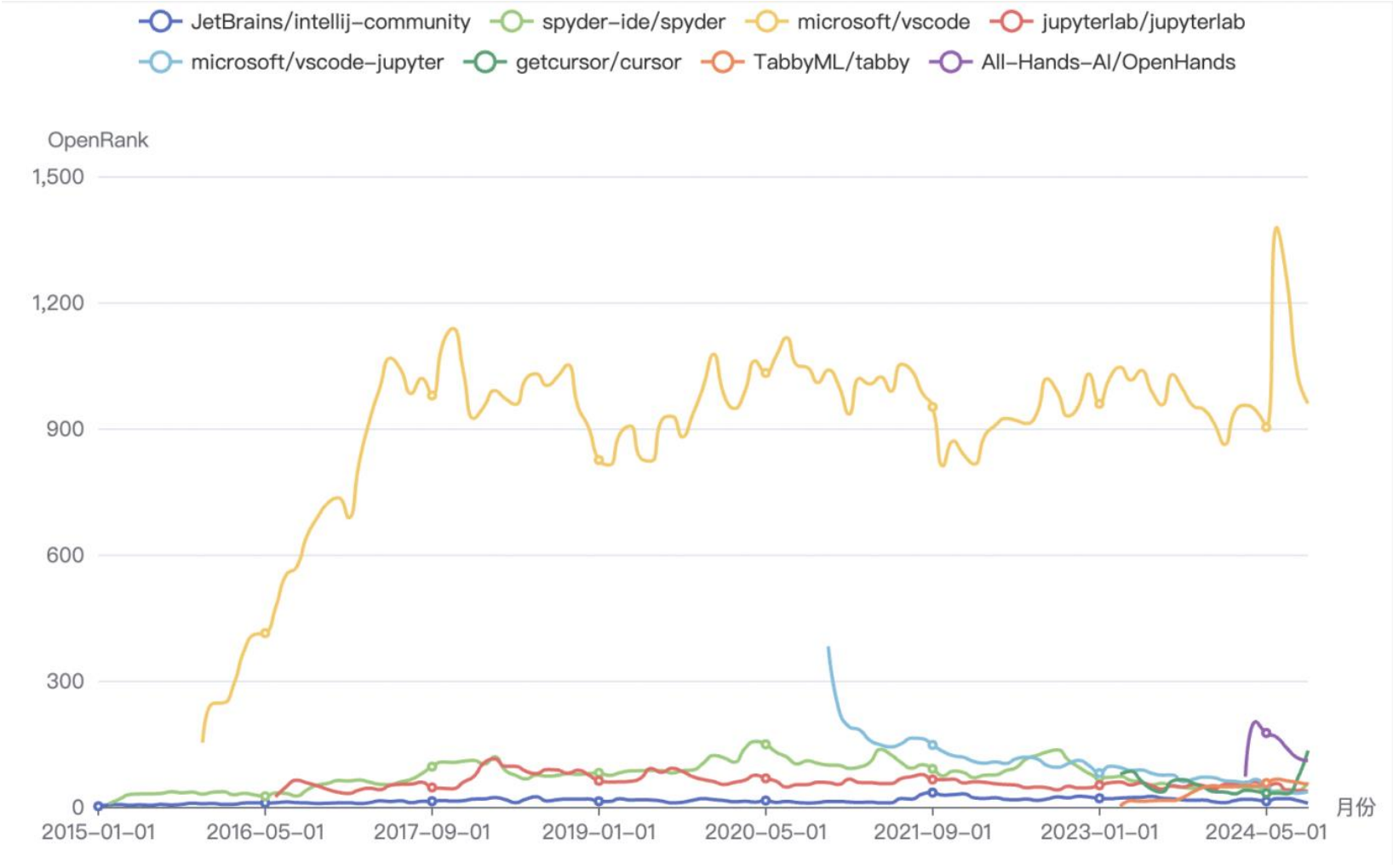
AI应用编排类框架



https://xiaoya-yaya.github.io/ant-ospo-insights/

传统IDE与AI交互式开发工具

ANT OPEN SOURCE



https://xiaoya-yaya.github.io/ant-ospo-insights/

04

回到架构

面对 AI 生态的全景图

AI生态中的生态位

ANT OPEN SOURCE

我们能做些什么

ANT OPEN SOURCE

我们好像看到了20年前的LAMP

- （给年轻朋友）LAMP 即 Linux，Apache httpd，MySQL，PHP；
- 大模型开发的范式正在形成，做应用，或是做 LAMP （平台、基础设施）本身；
- 与 LAMP 不完全相同的可能是，大概没有哪个组件会一统江湖，但生态是有些神似的。

对于要进入 AI 生态的参与者

- 关注趋势上的项目，和新的生态位，做自己的判断；
- 对于做自己的项目的人，最好能利用上一个好的生态位，或是和中心生态位产生良好互动；
- 如果有余力，创造新的生态位

QCon
全球软件开发大会

THANKS

大模型正在重新定义软件

Large Language Model Is Redefining
The Software

为开源世界带来微小而美好的改变