

范昊

17736376130 | jorybeast97@foxmail.com | 北京

1997-09 | 男

在职 | 后端开发

教育经历

华北理工大学

2017年09月 - 2021年06月

电子信息科学与技术 本科 人工智能学院

专业技能

Go / Java / Python; 分布式系统、事件驱动、MQ、异步任务编排、并发控制、Docker / Sandbox、SSE、网络代理与长任务恢复; Agent Harness、ReAct / Tool Calling

工作经历

小红书

2026年07月 - 至今

Rednote国际化后端研发组

负责海外TnS (Turst & Safety) 国际化架构搭建及迁移, 建设贴合Rednote One App战略的AI Native审核系统。

北京字节跳动科技有限公司

2024年07月 - 2026年07月

TikTok Ecosystem Platform - Backend Engineer (L3)

Vinbo Agent

背景: 面向内容治理运营建设的 Web Agent 与通用 Harness。运营可以通过自然语言调用已接入的 Skill / Tool, 串联风险识别、模型训练和线上处置, 完成端到端治理。Harness 与执行环境、工具注册和异步任务编排解耦, 同时通过 API 为其他业务提供长任务接入能力; Vine Plus 的原子风险能力以 Skill / Tool 形式供 Vinbo 调用, 两个平台通过任务关联形成治理闭环。

- 负责 Vinbo Agent Harness 的整体架构设计与工程落地, 覆盖图形化 Client 与独立 Web 执行端; 以 ReAct / Tool Calling 组织多轮推理, 统一接入 15+ 治理工具、Skills 与子 Agent。
- 推动 Web 端长任务从 Pod 内进程循环演进为事件驱动执行, 围绕事件持久化、模型 Chunk 聚合、异步编排和服务重启后的任务接管, 支撑训练、回捞、打标等任务持续推进。
- 建设 Agent Monitor, 并接入 Fornax Trace, 会话关联模型请求、工具执行和任务结果, 支持异常发现、批量诊断、定时巡检和 LLM 辅助归因。

Vine Plus Platform

背景: 面向内容理解的 AI 平权平台, 让运营通过点选配置和少样本训练获得风险识别能力, 并基于评估集持续迭代 (active learning)。平台同时提供原子模型能力和 Workflow: QuickModel 以 Embedding 聚合内容特征、分类头完成风险分类, 配合文本模型、敏感词匹配、SFT、PE和相似召回等能力, 最终承接 TikTok 发文、内容变更等流量, 经过模型和策略判断后输出处罚或处置结果。

- QuickModel 与 Active Learning: 负责从 Labeling 生产、数据存储、训练任务、Checkpoint 产物, 到在线服务部署和特征生产的完整链路; 支持模型多轮训练与样本回流, 打通 AI 标注替代人工标注后的持续迭代流程; 针对每轮 AI 标注产生的正负样本, 结合样本难度、相似性和分布选择进入下一轮训练集, 避免训练数据被简单样本或过度集中的困难样本主导, 提升 Active Learning 的有效性。
- Worklow 治理流: 负责基于事件驱动的端到端 Workflow, 处理上游召回流量的特征拆分与合并; 策略迭代时复制部分流量执行 try-run, 并与线上结果进行 case-level diff, 为策略验证和回归分析提供依据。

北京字节跳动科技有限公司

2022年04月 - 2024年07月

抖音 内容生态

北京

Ark内容治理平台

背景: 闭环内容生态治理全流程, 包含发现-标准-识别-机/人审-处置-感知等环节, 提供一站式的内容治理能力, 支持业务定制化及二次开发能力, 覆盖视频、图文、商业化、抖极/抖火等多业务场景, 提升整体生态问题治理效率。

工作内容:

- 生态特征优化: 主导生态特征读取链路开发, 峰值QPS达270k+, 利用一致性哈希、singleflight及localCache

ARC策略构建多级缓存体系，头部特征缓存命中率提升至43%，集群整体负载降低32%，通过DAG依赖优化，系统并发能力显著提升，P99延迟控制在50ms以内。

2. 动态审核优化：负责人审链路的优化，实施削峰填谷策略，结合动态调控预测，整合模型及风险信号等多维度特征，实现风险梯度治理方案（全审、抽审、机审、放出），全面覆盖生态审核场景，使整体人效提升46%，审出率提高23%，审核延时趋近于零。

生态处置平台

背景：收敛生态所有处置能力，提供统一接入并规范化接入流程，降低接入成本并提升整体可用性和灵活性。

工作内容：

1. 平台设计&落地：独立负责架构设计与全链路落地（SDK、网关、增强服务、核心处置、后置适配），采用事件驱动架构，支撑 QPS 1.4w+、SLA 99.995%；多次下游故障中实现快速回溯，保障处置最终一致性。
2. 多租户能力扩展：扩展多租户与独立集群部署能力，支持外部业务独立接入与二次开发，提升资源隔离与定制化能力。

抖音原创专项

背景：治理侧原创专项，打击抄袭/搬运，提升原创内容占比与平台调性。

工作内容：

1. 原创识别体系搭建：负责原创项目 V2.0 标签审核体系，搭建多环节原子识别与碰撞判责链路，日均识别非原创内容 540w+。
2. 推荐召回优化：优化原创内容推荐召回，经 Libra A/B 验证，在提升治理效果的同时带动用户留存提升。

百度在线网络技术公司

2020年12月 - 2022年04月

后端开发工程师 搜索策略部 效率平台

北京

背景：Debug平台用于对线上 & 调研环境问题case进行快速定位，并打通模型在线分析服务，提升检索效果问题解决效率。

工作内容：

1. Debug Case分析能力建设：开发一站式CASE分析功能，支持日志Trace & 实时分析，采用 算子编排 + FaaS化实现用户自助分析，实现检索结果丢失、逆序等Case各阶段分析，排序模型在线计算及DEBUG功能，Case分析成本15min+降低到 30s内大幅降低问题分析成本。
2. 抓取平台：优化抓取平台任务调度流程，采用EDA架构、汇聚流式处理结果等方式，提升整体抓取任务吞吐量，降低整体抓取集群负载。