

何星泽

15022649329 | hxz2046084122@outlook.com

工作经历

蓝色光标 DSP 广告平台

算法工程师 | 2025.07 - 至今

核心模块一：DSP 投放引擎架构设计

设计并落地 DSP 投放引擎全链路架构，支撑日均百亿级 RTB 竞价请求（常态 QPS 50w+），覆盖从流量接入、竞价决策、出价计算到归因分析的完整链路。

- 基于 Go 构建高性能竞价引擎（Bidmaster Engine），实现多 ADX 协议适配与实时 RTB 请求解析，竞价响应 P99 < 100 ms。
- 设计模块化策略管线，将定向过滤、频控、出价等决策逻辑解耦为独立插件，支持配置热加载与灰度发布，策略上线周期从天级缩短至小时级。
- 构建基于 Kafka 的异步日志管线与实时竞价数据采集体系，为下游模型训练、策略迭代和漏斗分析提供数据基础。

核心模块二：智能出价与策略优化体系

针对广告主成本控制（CPA/CPC）严格、平台预算消耗非线性（利用率仅 85%）及新计划冷启动困难等核心痛点，搭建智能出价策略体系：

- 研发 PID 控制算法与非线性减速策略，基于时间片动态分配预算，实现分钟级数据轮询热更新，平滑全天消耗曲线。
- 设计基于 K 值的自适应出价系数（Smart Bidding），针对 CPC/OCPC 场景动态调整出价；引入慢启动策略（Slow Start）保护新计划冷启动。
- 建立基于后验 CTR 的校准与兜底逻辑，确保数据稀疏时期的出价安全性。

落地效果：OCPC 成本达标率 68% → 85%，平均成本偏差降低 68%；预算利用率提升至 98% (+15.3%)，整体 ROI 提升

核心模块三：DSP-Model 广告预估模型体系

全权负责 CTR（点击率）、CVR（转化率）、WINR（竞价胜率）三类预估模型的研发与迭代，直接影响平台竞价排序与广告主成本控制。

- 基于 TensorFlow + Horovod 搭建分布式训练框架，采用 TFRA 动态 Embedding 处理大规模稀疏特征，支持千万级样本训练。
- 设计多任务学习架构（MTL）：CVR 模型支持激活/付费/注册/唤醒多目标联合建模，通过 Shared Embedding + Task-Specific FM/DNN Heads 实现特征复用与目标解耦。
- 引入保序校准层（Isotonic Regression），校准预估概率与真实转化率偏差，校准后 PCOC 趋近 1.0。
- 建立模型全生命周期管理：自动化调度（join/update 两阶段）→ 离线评估（AUC/GAUC/Logloss/PCOC）→ SavedModel 导出 → 线上 A/B → 飞书告警。

落地效果：多目标 CVR 模型核心任务离线 AUC 达到 0.80+，模型迭代周期从周级缩短至天级。

核心模块四：AB 实验平台与基础设施

算法团队此前缺乏高效的并行验证手段，模型与策略验证周期长达 2 周。从 0 到 1 设计并落地支持多层正交的 AB 实验系统：

- 设计 CTR/CVR/策略三层正交分流架构，实现基于 UserID/DeviceID 的一致性哈希分流，分流耗时 <1ms。
- 开发配置热加载机制（30 秒生效），构建基于 BOS/COS 的多云数据导出与可视化漏斗分析系统。
- 建立全链路日志埋点（请求 → 竞价 → 展示 → 点击 → 转化），支持小时级指标自动聚合。

快手 - 北京快手科技有限公司

IT中心 后端研发实习生 | 2025.01 - 2025.05

- 参与 Kit 办公融合产品的设备监控系统优化，基于 Redis 实现设备状态跟踪与健康评分算法，扩展 CPU/内存/网络延迟等多维监控指标。
- 负责 KRooms 会议室系统的 API 性能优化与认证模块修复，将接口响应时间从 120ms 降至 40ms，API 错误率从 2.5% 降至 0.5%。

项目经历

AIMO-Tech | 联合创始人 & 技术负责人

2025.11 - 至今

负责 AIMO-Tech 产品矩阵的核心技术架构设计与全栈开发，涵盖 **AI Agent 框架**、**LLM 服务基础设施**与开发者平台。

AgentKit 执行引擎 (Go)：设计并实现通用 Agent 执行框架，支持工具调用编排（Tool Executor + 多步推理链）、会话状态管理（Session/Context Repo）、流式事件驱动架构（Event Emitter），以及虚拟文件系统挂载（VFS），使 Agent 能通过标准文件操作（ls/view）访问外部业务数据。为上层 Agent 应用（智能问答、素材管理、品牌知识库）提供统一运行时。

DotAI 多模态 Agent 系统 (Java)：构建面向企业场景的 AI Agent 产品，支持品牌知识库检索、素材管理与多轮对话。设计 Virtual File System 映射层，将 Agent 工具调用转换为对业务 API 的标准化访问，实现 Agent 与企业数据系统的松耦合集成。

竞赛与开源

Apache Higress RAG 赛道 | 决赛入围：搭建"Pre-Retrieval → 多路混合检索 → Post-Retrieval → CRAG 纠错闭环"全链路增强架构，覆盖 Query Rewrite / Rerank / 压缩 / 纠错检索等模块。设计多路检索与路由策略（向量 + BM25 + WebSearch），支持 RRF 融合与意图路由。实现 CRAG 纠错检索闭环，在 HotpotQA 数据集上达到 **EM=0.94**（10,000 样本）。

TAAC 搜广推大模型竞赛 | 参赛中：参与搜广推场景下的大模型 for Recommendation 方向竞赛，探索 LLM 在用户行为建模与推荐排序中的应用，当前 Baseline **AUC=0.8486**。

开源贡献：Apache Higress（云原生 API 网关）、seata-go（分布式事务框架）贡献者，对云原生与大模型应用落地有深入研究。

专业技能

编程语言：精通 Python、Go，熟悉 TypeScript、Java；

搜广推算法：CTR/CVR/WINR 预估模型、智能出价（PID / Smart Bidding / OCPC）、预算控制与消耗平滑、AB 实验设计与分析；

机器学习工程：TensorFlow + Horovod 分布式训练、TFRA 动态 Embedding、多任务学习（MTL）、保序校准、模型全生命周期管理；

Agent & LLM：Agent 执行引擎设计与 Tool-use 编排、多模态 Agent、虚拟文件系统、RAG 全链路（向量检索 + BM25 + CRAG）、LLM 服务可用性监测；

系统架构：高性能 RTB 竞价引擎（QPS 50w+）、微服务架构、Kafka 消息管线、Redis 缓存策略、分布式一致性（Paxos/Raft）

教育经历

天津财经大学 - 本科

信息与计算科学

2021.09 - 2025.06