

公域数据降噪与增量优化

2025-12-03 17:50 - 雷文武

状态:	开发中	开始日期:	2025-12-03
优先级:	普通	计划完成日期:	2025-12-19
指派给:	雷文武	% 完成:	0%
类别:		预期时间:	0.00 小时
目标版本:		耗时:	32.00 小时
等级:	简单	版本:	
变更任务:			

描述

*e火山侧每天采集约120w 条公域数据

*e模型实际成功打标的数量仅10w+

*e假设模型本身无问题，则：

- 至少80% 为噪音/广告/业务无关内容
- 需从采集入口与数据清洗环节减少无效数据，降低模型GPU使用资源

历史记录

#1 - 2025-12-03 17:50 - 雷文武

- 状态 从 待处理 变更为 开发中