

资料阅读说明

开发回归记录。自动化通过数、模型样本结果和电脑JVM开销不等于真实总体准确率、手机全流程速度或多厂商通知可靠性。

本地解析与提醒意图验证

日期：2026-10-12。对应Android v0.8.0-dev、规则库0.6.0及网关协议2。

1. 本轮目标

将用户指定的提醒时刻与事项发生时刻分开；让可确定的短句和单字段补充在本地结束，减少模型请求。复杂通知和图片继续调用模型，随后经过相同的依据、时间与提醒校验。

输入	预期
明天下午3点提醒我参加学校会议	次日15:00提醒
明天下午3点参加学校会议	次日14:30提醒，事项15:00开始
明天下午3点参加会议，提前10分钟提醒我	次日14:50提醒
下午3点提醒我打印学期注册标签	首次提交当天15:00；已过则确认日期
中午3点提醒我打印学期注册标签	本地只确认是否指下午3点；不问消息日期
明天九点半提醒我取票	本地询问早上或晚上，选项保留半点
半小时后提醒我取书	从首次提交倒计时，重试不重新计时
报名明晚20:00截止	直接输入按次日19:00提醒；粘贴且消息日期不明时先确认消息日期
找老师签字	无提醒待办，不追问时间

“中午3点”本身存在矛盾，仍须确认一次具体时刻。短句引用、条件、否定、多节点、重复和改期不会因长度短而强行本地放行。

2. 自动化覆盖

- 时间：中文/数字钟点、半点、一刻/三刻、12点、凌晨0点、跨午夜、相对星期、完整日期、非法日期、过去日期、只有日期、时区边界。
- 提醒：指定时刻、明确提前量、准时、开始默认提前30分钟、截止默认提前1小时、开放与截止窗口、结束不提醒、无时间待办、提醒晚于截止的冲突、相同触发点去重。
- 补充：早晚、矛盾钟点、事项日期、消息日期、年份、缺失钟点、分钟保留；同一事项完成本地补充且不需要模型配置。
- 稳定性：连点、页面重建、SQLite迁移、通知答案过期、旧模型响应、删除/完成/手动修改优先、任务登记失败、有界重试、原文及回答保留。
- 产品功能：笔记独立保存、分享、批量操作、详情与提醒结果页、通知处理、系统权限和生命周期。
- 网关：原文依据、结构错误与截断、协议兼容、提前量依据、上游鉴权/额度/过载/网络错误、请求鉴权。

测试源位于app/src/test/java/com/shishi/app与server/*.test.mjs；文字和图片开发样本位于evaluation/reminder-intent-cases.json与evaluation/clarification-stability-cases.json。图片均为合成通知，没有使用真实群聊截图。

3. 执行方式与证据

本地测试通过scripts/build.ps1执行assembleDebug、testDebugUnitTest与lintDebug；服务测试使用node --test server/*.test.mjs。

云端评测将合成输入发送到实际部署的网关，检查结构化事实并捕获结果。LiveAdoptionTest随后将这些真实响应交给Android实际使用的ModelWorkflow、时间校验及ReminderPlanner，检查最终状态、提醒数量和准确触发时刻；不把模型返回200当作完成验收。模型不能识别的“提醒晚于截止”关系由本地规划器阻止调度。

首次图片回归发现网关只接受截短的提前指令，拒绝模型引用的完整原句。已允许完整连续原文中唯一且数值一致的提前指令，并加入服务回归；多条提前要求、错误数值和编造依据仍拒绝。

原始模型响应、云端部署元数据和临时结果仅保存在Git忽略的.local中；本文保留脱敏结果与复现说明，不提交密钥。

4. 最终结果

检查	结果
Android单元及运行模拟	171项通过，0失败、0跳过
网关与迁移测试	24项通过，0失败
APK构建与Lint	通过；Lint无问题
真实云端开发回归	40/40通过：新协议24、旧协议16；29文字输入、11次图片输入，图片为7张独立合成图
真实响应在Android采用	40/40通过；检查最终状态、数量和精确触发时刻
本地核心性能	9类输入预热450次，测量900次；p50 0.1722ms、p95 0.3063ms、均值0.1786ms；均无需AI分流
APK	versionCode 15；0.8.0-dev；997226字节；签名与上一包一致
云端	shishi-parser-022为稳定版本，100%流量；最小副本0、最大1，原秘密配置保持；健康200，未授权401

APK SHA-256：a38ca50db6c459c3b1ce8d41073603bc914045cbbf6c1c130e0120c457d0b759。

一次云端回归仍有秒级耗时波动，不能承诺复杂通知或图片即时完成。效率改善来自常见短句和必要补充不再发送模型请求；本地性能数字仅为此次电脑JVM测量。

5. 验证边界

开发集用于回归和定位问题，不代表总体识别准确率。电脑JVM本地耗时不代表手机端完整流程或能耗。Android模拟使用Robolectric API 34，尚未完成Android 16真机验收、多厂商后台通知到达和独立保留集评估。

本轮不新增常驻后台服务或轮询。复杂通知仍可能等待网络及模型；重复系列完整调度、部分节点先提醒、已有历史事项自动改期匹配仍未实现。