会议纪要｜AI 接口技术细节 & 问题排查

会议主题：AI 接口的技术细节、调用方法与问题排查

结论关键词：优先使用流式接口｜空参数采用“删除字段”策略｜502 报错需网关/后端联合排查｜统一 SID 处理规范

一、接口类型与调用方式（达成共识）

1. 非流式（Text Chat）：一次性返回完整结果，适合短文本或无需渐进式反馈的场景。

2. 流式（Streaming）：以数据流逐步返回，官方推荐优先，可提升首字节时间（TTFB）与交互体验。

3. 调试方式：通过 “check it out”等测试面板或 Postman/cURL 直接发起请求；message 参数可传入 “hello” 等进行快速联通性测试。

决议：当前阶段默认走流式接口，非流式作为回退方案。

二、已识别问题与影响

1. 空值参数（SID）——传入空的 Session ID 导致接口报错，增加异常率。

2. 参数处理分歧——争议点是传递空值还是直接删除字段，风险在于严格校验可能触发错误。

3. 稳定性（502 Bad Gateway）——流式接口调用出现 502，需定位网关转发或后端可用性问题。

三、统一规范（当场决定）

1. 空参数处理：一律不传空值字段，仅当有值时才下发。

2. 错误重试与降级：流式失败后回退一次非流式，再失败提示重试并打点记录。

3. 日志最小集：需包含 requestId/traceId、路由名、时间戳、响应码、耗时、实例名称。

四、待确认事项

“刘氏接口”需确认其具体功能定位与依赖关系，由 Ch 会后澄清并同步文档。

五、行动项（Action Items）

[zle] 收集并回传 502 报错日志，含 requestId/traceId。验证删除空字段策略。

[Ch] 排查 502 根因，输出《刘氏接口功能说明》。

[双方] 制定 SID 处理规范示例，建立冒烟用例作为回归基线。

六、风险与缓解

1. 接口不一致：需统一封装规避。

2. 流式链路不稳定：需配置心跳与缓冲机制。

3. 观测不足：需补齐端到端 trace 与监控指标。

七、里程碑（建议）

T+0：确定 SID 策略并合入；补充日志字段。

T+1：完成 502 初步定位，提出修复建议。

T+2：回归测试并生成稳定性周报。

附录：请求体与异常处理建议（示例）

A. 请求体构造示例（JavaScript/Pseudo）：

const payload = {  
 messages,  
 ...(sid ? { session\_id: sid } : {})  
};

B. 错误处理流程：  
调用流式 → 502/超时 → 记录traceId → 自动回退非流式一次 → 若失败提示重试并上报。

C. 冒烟用例：  
1) 流式+SID；2) 流式无SID；3) 非流式无SID；4) 异常注入触发降级。