六分钟实验量化结果

第一视角影片 00:00–06:00 | 现场任务：拆开落地风扇进行清洁

6 / 7
得到文字答复的请求

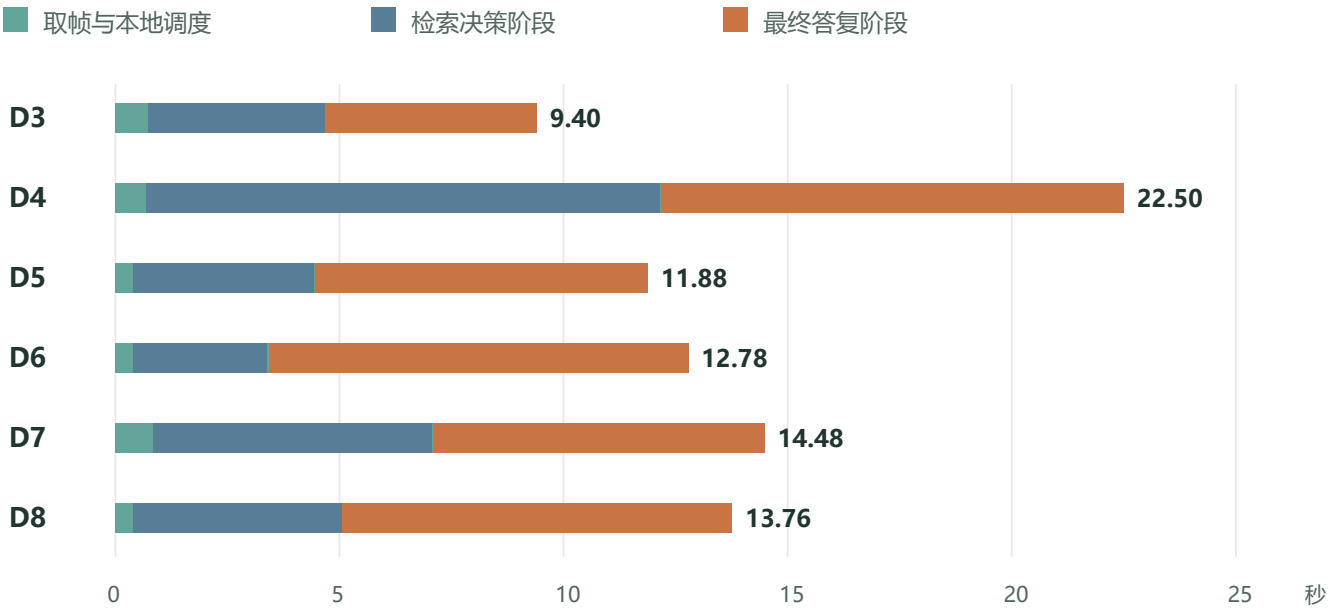
13.27 秒
请求至文字答复 · 中位数

13.46 秒
请求至开始播报 · 中位数

7 次请求中有 1 次因任务或指令变化而撤销。6 条答复含 1 条澄清，均有播报完成回调。

每轮等待文字答复的时间

每条从系统创建请求开始计时，到文字答复生成结束；只计 6 个完成回合。



两次模型调用阶段占累计文字等待时间的 95.8%。
两类耗时均包含传输和服务端处理。本次 6 轮均未实际执行网页检索。

画面能核验多少回答?

5 / 6 条答复包含可用当轮图像核查的主要肯定陈述。

5 / 5 条可评陈述与对应画面一致。另 1 条澄清未作同类陈述。

这是对六条回复的 AI 辅助初审。整套拆装指导的正确率与任务完成率仍缺独立人工真值。

数据来源 · run-1795163319261 的 events.jsonl、当轮输入 JPEG 与眼镜播报回调。
计时起点在语音识别之后。现场说完至系统建请求的完整时延没有测得。