

无人值守验收能力说明书

本说明书面向把工作交给流马的业务负责人和验收责任人。它说明如何设定判据、如何看过程证据、何时可以接收结果，以及何时必须停止并找管理员。本文不是安装手册。

1. 这项能力适合什么工作

适合以下任务：结果需要多轮尝试、验收标准可以写成明确规则或脚本、用户不想在每一轮旁边盯着，但又不能只接受“模型说完成了”。

典型用法是：

1. 用一句话说明要完成的工作和不可违反的约束。
2. 如有自己的验收脚本，把它作为附件一起交给流马，并明确“脚本不许改”。
3. 流马回显目标、验收标准和预算。
4. 流马自主产生候选、运行验收器、读取失败证据并继续改进。
5. 每轮结果写入判定卡；交付前先渲染；最后才发交付卡。

这项能力不替代最终业务验收，也不保证一次首页渲染能覆盖所有页面。

2. 使用前提

- 管理员已经为组织启用 `pursue system skill`。
- 任务使用的数据和文件已经放在当前组织工作区；本演示使用 `base/demo-base.json`。
- 验收器能在允许的执行环境中运行，不依赖未配置的凭证或外部系统。
- 验收标准必须能被明确表达。无法验证的“好看一点”“更专业”等要求，需要先改成可观察结果。
- 涉及生产写入、客户数据、凭证或对外发送时，必须另行确认权限；本演示不涉及这些动作。

本能力本身零安装、零外部连接。管理员侧的平台部署不属于本说明书范围。

3. 日常怎么说

交活

在会话中输入 `/pursue`，随后说明目标、硬约束、需要读取的材料以及交付前必须执行的检查。例如：

`/pursue` 读取工作区里的数据底座，按其中数据生成一个单文件内部系统。附件是我的验收脚本；脚本不许改，必须跑到 `verdict=pass`；交付前请实际渲染结果。

如果验收器是买方自己的，请把所有权边界写明：流马只能运行它，不能修改它。

修改判据

直接在同一会话里说出新增或修改的标准，不需要进入审批页面。例如：

验收标准再加两条：看板必须显示全部记录的金额合计；详情页必须显示该对象关联的全部记录。

本次实况中没有生成第二张契约卡，因此修改后要同时检查用户消息、流马回应和下一轮判定卡，确认新条件确实进入了候选和判定说明。

4. 怎么检查流马理解得对不对

收到第一张“Pursue 验收契约”后，检查三部分：

契约部分	要检查什么
目标引用	交付对象、文件或运行目标是否正确
验收标准	是否覆盖你的硬约束；验收器命令和判据名称是否正确
预算	最大轮数、最长时间和停滞阈值是否符合预期

不同意时立即在会话中改，不要等到交付后再纠正。

注意：当前产品会回显预算，但本次现场没有发现运行时读取或强制预算的证据。预算卡不能当作已经生效的硬停止器。

5. 怎么读判定卡

每张判定卡至少要回答：

- 本轮是 **pass** 还是 **fail**；
- 得分是多少；
- 哪些具名判据分别通过或失败；
- 环境证据引用在哪里；
- 本轮候选与上一轮相比改了什么。

本演示的四条判据是：

判据	含义
single_file	只有一个 HTML，零外部依赖
reference_integrity	页面、菜单、返回和下钻引用都指向已知页面
positive_connectivity	从入口页可以到达所有页面
data_consistency	带 derive 的指标可从数据源重算且数值一致

pass 只证明这些判据覆盖的范围通过。它不自动证明脚本没有检查的运行时行为。

6. 什么时候可以接收交付

至少同时满足以下条件：

1. 冻结验收器的最终结果为 **verdict=pass**。
2. 判定卡逐项列出具名判据及证据引用，而不是只有一句“全部通过”。
3. 交付前已经产生真实渲染截图。
4. 用户后续追加的标准已经体现在最终结果和最后一轮判定中。
5. 交付卡出现在最终判定之后。
6. 对关键交互做只读实走；不能只看首页。

本演示最终可复核的数字为：运单 8、司机 5、已送达 5、运输中 2、已取消 1、已送达运费 **12800**、全部运单运费 **17000**。

7. 自动运行时会有什么

启动 `/pursue` 后，流马会在同一会话中持续推进候选，不需要用户逐轮点击“继续”。用户可以离开，回来后按时间顺序查看：契约卡、工具证据、渲染图片、判定卡和最终交付卡。

本能力没有在本次现场验证定时启动、外部通知或主动告警；不要把这些行为当成已验证能力。

8. 常见失败与处理

验收器返回 fail

这是正常收敛过程。检查失败项是否明确、是否有环境证据、下一轮是否针对该项修改。不要让流马通过改验收脚本来消除失败。

只有总述，没有具名分项

停止接收。要求判定卡列出每条判据和证据引用；否则无法判断究竟检查了什么。

脚本通过，但页面没有被看过

停止接收。要求先渲染，并对脚本明确不覆盖的页面或交互做只读实走。

页面部分字段为空

不要用脚本 `pass` 覆盖运行时缺陷。本演示的司机详情页就是反例：关联运单子表正确，但司机本人五个字段恒空。发现此类问题时停止交付，交给制造或修复阶段处理，不要在记录阶段补造证据。

run 已完成后仍出现新的迭代

找管理员核对账本。本次现场中，第二次 `mark_done` 被拒绝，但终态 `run` 仍接受 `log_iteration / select_best`，导致完成时间早于 `best iteration`。

HTML 无法正常读取

当前 `read_file` 对 HTML 存在已知问题，可能返回 Markdown 提取或极短内容。需要管理员确认读取路径；不要把空结果当作文件为空。

9. 数据与安全边界

- 业务数据属于当前组织；不要把真实客户值写入跨客户流通的 `trajectory` 母版。
- 凭证、`token`、`cookie`、`app id` 和生产路径不得进入截图、文件名或说明书。
- 文件系统中的 `skill`、`Recipe` 和身份定义不能在记录过程中修改。
- 记录阶段只允许浏览、只读查询、展开折叠和纯前端预览操作。
- 对外发送、发布或交付给客户需要另行授权；生成内容不等于获得发送权限。

本演示的公司、司机、运单、发货方、金额和日期全部为合成数据，无凭证、无 `token`、无外部平台。

10. 已知限制

1. 验收器只检查它声明的范围；运行时渲染和交互需要额外只读实走。
2. 首页自看不能替代全页面覆盖。司机详情页字段恒空正是本次漏检实例。
3. 终态 `run` 仍可写入部分账本动作，完成时间与 `best iteration` 顺序可能异常。
4. `read_file` 读取 HTML 不稳定。

5. `write_file` 保存底座时可能丢失末尾换行。
6. 契约预算当前只看到回显，没有看到运行时强制执行。
7. 非 Recipe 目标仍会收到 Recipe 专属 tick 指令，由流马自行判定为不适用。

11. 本次能力证据

- 演示会话：`e91ce597-2d8d-4ca4-b0d6-cca477550c9c`
- PursueRun：`e78f5a91-ee4d-4eef-aae5-1c666f8c6658`
- 最佳 iteration：`b45da0fd...`
- 交付物 sha256：`af063c01edc2845bea3156bad39602cdb3d08383fe2aaab2a4608024c6f71ad8`
- 完整图文流程：[demo-walkthrough.md](#)
- 真实现场、数字链和审计记录：[scene.md](#)

12. 接收原则

可以信任的是“明确判据 + 环境证据 + 实际渲染 + 关键路径实走”组成的证据链，而不是任何单独一项。脚本 pass、模型自述或一张首页截图都不够。