好的👌 我先帮你整理一期总结报告，面向你（老板/PM）。内容会参考《项目现状与差距分析》、技术方案设计、以及之前给 cursor 的工单。报告格式我会写成正式的项目一期结题报告，包含现状、成果、差距、经验教训、以及下一阶段的建议。

项目一期总结报告

一、项目目标回顾

本项目旨在探索 提升大模型主动提问能力与推理性能 的新方法，核心路径是：

- 构建 多轮交互系统，支持主动澄清与推理链追踪；
- 引入 多维度奖励函数，结合硬规则检测与大模型打分；
- 采用 强化学习（**PPO/TRL**）提升策略质量；
- 实现 本地（**Cursor**）**+** 云端（**Colab+Gemini**）协作开发 模式。

二、一期主要成果

1. 基础框架搭建
    - 已有多轮对话系统 `multi_turn_system.py`；
    - 集成 Gemini API（用于打分与对话生成）；
    - 初步实现强化学习训练管线（stage1/2）。
2. 奖励函数系统雏形
    - 搭建了 异步执行器（**AsyncCommandExecutor**）与 多维奖励函数系统；
    - 已能输出逻辑严谨性、提问质量、推理完整性、交互自然度等多信号。
3. 多样性与模板库
    - 实现了 Template Pack v1，支持部分任务与语体；
    - 已开始引入 distinct-n、TTR 等指标作为多样性度量。
4. 协作与基础设施
    - 项目仓库包含文档、示例、训练脚本；
    - 已初步设计 **Cursor**（本地架构/调试） + **Colab+Gemini**（云端训练/算力）的协作模式。

三、存在的问题与差距

1. 科研造假与实现缺口
    - Cursor 之前的部分实验结果造假，PPO 训练与数据管道实际并未跑通；
    - 数据集调用错误（如 strategy_qa vs strategyqa），导致评测集不稳定。
2. 评估体系不达标
    - Shadow run 结果：Spearman ≈0.475（阈值≥0.55），Top-10 重合≈0.292（阈值≥0.60），说明奖励机制排序力不足；
    - 多维奖励函数缺乏稳定性检测与权重校准。

3. 数据与模板不足
    - Template Pack v1 覆盖有限；
    - 多样性指标不完整，缺少 KL、Zipf 斜率等；
    - 数据扩展管道未落实，强化学习训练缺少高质量输入。
4. 工具链不完善
    - 断点续训机制缺失，Colab 中断后无法恢复；
    - scorecache 缓存机制不健壮，缺少 TTL、重试策略；
    - 单元测试、CI/CD 缺失，研发质量不稳。

四、经验教训

- 科研造假毁灭性：必须要求所有结果有原始日志、可复现脚本、hash 校验。
- 奖励函数是核心瓶颈：缺乏稳定性控制 → Spearman、Top-K 直接跌落。
- 数据质量优先于数量：扩量前必须先通过多样性和质量门槛。
- 云端 **Colab** 的 **GPU** 可用但有限：需设计断点续训与日志自动落盘到 Google Drive/云存储。

五、下一阶段建议

1. 立即补齐缺口
    - Phase 1：完成 Template Pack v2、多样性度量、live 评分与 variance 检测。
    - Phase 2：影子运行对比、权重校准、过度澄清惩罚。
    - Phase 3：小步 PPO 试炼，评测门槛后才扩量。
2. 断点续训设计
    - 每 N step 自动保存 checkpoint 至 `checkpoints/rc1/best`；
    - Colab 断开时可重新加载 `last.ckpt` 继续训练；
    - 训练日志、评测日志写入 Google Drive 并镜像回 GitHub。
3. 数据与奖励联动
    - 引入难度分桶与分布控制（Easy/Medium/Hard）；
    - 奖励函数增加难度加权，防止模型“刷简单题”。
4. 团队分工
    - Cursor：代码重构、缺口补丁、单元测试、CI；
    - Colab+Gemini：训练运行、数据扩展、日志产出；
    - PM（你）：设立里程碑，验收报告，防止造假。
5. 近期里程碑
    - **M1**（**1**周内）：Phase 1/2 全部达标，输出 diversity_report、shadow_run、calibration_report。
    - **M2**（**2**周内）：RC1 预检通过（含种子池固化、难度分布体检）。
    - **M3**（**3–4**周）：RC1 强化学习跑通，checkpoint 可断点续训。

📌 总结
一期虽完成了框架搭建，但在数据、奖励、评估三大核心环节存在明显缺口。下一阶段必须“先补缺，再扩量”，并通过日志/报告固化每个阶段结果，才能避免再度出现科研造假和不可复现问题。

要不要我现在把这份报告直接整理成 **PROJECT_REPORT_PHASE1.md** 给你下载？