六目拼接项目 · 总控台

实验一点评与总评

八组实验一点评与作业项目总评

文档编号	ZKT-2026-DOC-05
评审基准	任务书 11/12 号 · 全班统一五步模板
评审对象	课程网站八组提交 · 含证据包
版　　本	v1.1（通知稿）

2026 年 11 月 6 日

八组实验一点评与项目总评

评审基准：任务书 11/12 号 • 材料截至 2026-11-06 16:30 截止

文档编号 ZKT-2026-DOC-05 • v1.1（通知稿）

1 评审基准与方法

实验一（大模型 Agent 模式配置及应用）的验收口径取自任务书 11/12 号：各组配置 GLM agent（Cline/ZCode），解读**本组角色对口**的代码对象，回答三问（数据从哪来 / 网络结构 / 输出是什么），AI 原话留档、逐条人工核对并标注对/错/存疑，**至少纠正一处 AI 错误**，交付证据三件套（命令全文、完整日志、产物路径）。全组通用纪律强调“证据包制”与“AI 可以用、结果自己扛”。

本次评审逐组精读了课程网站上的八份报告主体与证据包材料（报告文本全部提取复核），并与各组的项目角色（组 1 匹配决策、组 2 特征质量、组 3 缝合融合、组 4 数据与标签、组 5 测试与评测、组 6–8 端到端数据/训练/优化）逐一对照。评审回答一个核心问题：**各组是否达到了预期的项目理解**。

2 总体结论

1. **整体超出“配置应用”层面的预期：**八组**全部**对口解读了本组角色的代码对象（无一跑偏），八组**全部**达成“至少纠正一处 AI 错误”的硬指标，且多数组的纠错具有真实技术含量（见逐组点评）。证据包文化已经建立起来，“AI 解读 + 行号级人工对账”的方法论被普遍接受并执行。

2. **头部两组（组 2、组 3）达到可复用产出水平：**组 2识破了任务书提示词中埋设的两处代码陷阱；组 3 用三通道交叉验证 + 53 处行号机器对账产出 22 项写死参数清单，并发现一处真实规格矛盾（U1，见第 5 节）。

3. **理解深度普遍到位、工程规范参差：**理解层面八组均合格；但报告体例、证据完整性、署名与分工信息等工程规范差异较大（组 8报告为对话原文粘贴、组 5 姓名栏空白、组 1 无独立证据包）。

3 逐组点评

3.1 第一组（campus03 · 九头 · 匹配决策）——评级 B+

解读对象 `group04_mc/train.py`，与角色对口。**亮点：**三问答案全部落到行号级依据；最重要的纠错是指出 `p_correct` 并非已校准概率——末层为裸 `Linear(32,1)` 无 Sigmoid、损失为 MSE、标签为规则生成的伪标签，“输出概率”只是设计意图。这一纠正直指后续 mc 头验收（第一个精度验收头）的语义边界，理解深度到位。诚实标注了“torch 未安装导致动态验证未完成”。**不足：**无独立证据包（仅报告单文件），任务书要求的配置截图缺失；AI 原话为摘录而非

全量留档；报告在截止日 15:39 才上传，赶工痕迹明显。**建议**：补配置截图与完整对话记录；本机安装 torch 后补 `--smoke` 链路实测。

3.2 第二组（campus04 · 九头 · 特征质量）——评级 A+（标杆）

解读对象 `head_specs.py` 本组三头，完全对口。**亮点**：一页规格说明书达到工程文档水准（三头输入输出逐维解释、损失与优化器、每个结论带“文件: 行号”）；**识破任务书第一轮提示词埋设的两处陷阱**——temporal输入实为 $[B, 24]$ 无时间维（“seq[T,24]”只是数据来源描述）、输出为单个 $[B, 8]$ 拼接张量而非两个输出，并复核确认 AI 的反驳成立；人工另纠两处小错（35 行实为 34 行、一处引用错位）；证据三件套最完整（137 KB 完整会话记录 + 7 次工具调用命令全文 + 精确到秒的时间线）。**不足**：配置环节以表格代截图；6 人贡献度均为填平的 15–20%，分工颗粒度粗。**建议**：本组报告可作为后续实验的报告体例范本。

3.3 第三组（campus05 · 九头 · 缝合融合）——评级 A+（标杆）

解读对象 `stitch_initialize.cpp` 增益补偿段，即本组 gain_field 头未来的接管对象，选题意识最好。**亮点**：三通道交叉验证（Cline 插件两轮 + Cline CLI 两轮 + ZCode 独立解读），53 处行号引用经脚本机器对账 53/53 吻合；产出 22 项写死参数 + 6 项易混排除项的清单，直接服务 group07_gain 设计；**发现真实规格矛盾**：`learned_params.h:194` 声明 16 维而 `head_specs.py:28` 与 `onnx_provider.cpp:91` 均为 12 维（U1）；从 4080 经 SSH 只读下载源码，纪律执行到位；五人分工表明确到环节。**不足**：证据包目录层级偏深（双层嵌套同名目录）；两处行号纠错属“结论对、行号漂移”级别，无实质缺陷。

3.4 第四组（campus06 · 数据与标签）——评级 A-

解读对象 `label_builder.py` 三段打标 + 遥测样本，完全对口。**亮点**：五段式泳道数据流图（含任务单要求的校验闸门红框）；10 条 AI 解读逐条核对（对 7 / 错 1 / 存疑 2），纠正 p_correct 的望文生义式误读；系统揭示“同名不同义”风险（label_source 三段语义不同、params 同名两用）；两路线四维对照表落到字段级。**不足**：摘要笔误写成“实验四”；报告仅署名一人（其余组员列同组名单），分工信息弱于组 3/组 6；引用的遥测下载时间线里服务器写作“4090”应为 4080。**建议**：将表 1–3 的字段口径作为九头标签使用的边界依据归档（U7）。

3.5 第五组（campus07 · 测试与评测）——评级 B+

解读对象 `ab_compare.py` + `run_regression.ps1`，完全对口。**亮点**：七项指标的方向表（3 项性能 + 4 项一致性/几何）澄清了 init_ms 不摊帧等指定坑；91 处行号对账全部吻合；把“AI 在源码不存在时如实拒绝而非编造”（第 1–3 轮）记录为可信协作前提，方法论意识好。**不足**：报告主体仅两页、**姓名/学号栏空白未填**；无独立撰写的分工说明；遗留追问四项未闭环。**建议**：补署名与分工；遗留四项在实验二前闭环。

3.6 第六组（campus08 · 端到端 · 数据）——评级 A

解读对象 `datasets/README.md` + `make_demo_data.py`，完全对口。**亮点**：两轮解读 +“打架裁决”发现**任务单口径与代码事实不符**——生成时的平移真值 shift 并未落盘，训练回归目标实为加载时相位相关现场估计的伪真值（`train.py:64-67,77-82`），并以 SmoothL1 与张量契约佐证；发现 demo2 批次存在于磁盘但未登记进 README 的治理缺口；质量审计规范（17 处行号 sed 复核、计数修正留痕、失效命令版本一并存档）；五人各自独立跑一轮再交叉验证。**不足**：证据包归档于“组内”表述，网站附录的组织性一般。**建议**：U2 两项治理修正尽快入库。

3.7 第七组（campus09 · 端到端 · 训练）——评级 A

解读对象 `train.py` + `export_onnx.py` + `infer_test.py`，完全对口。**亮点**：八组中技术含金量最高的纠错——AI误报“原图非 512 时尺度不一致”，组员以**闭环代数证伪**（三步缩放净乘恒等于 1）+ 数值例 + 一万次随机闭环检验推翻，AI 复核后承认并反思成因；参数量独立复算 106,018 与 AI 对账，另以 checkpoint 字节数 434,309 B 交叉验证；发现 docstring“ResNet 风格”名不副实与 L32 转义瑕疵（U5）；4080 只读纪律明确。**不足**：报告主体偏简、仅署名提交人一人（组内分工未呈现）；数据流图为附图未入正文。**建议**：补组内分工；把闭环检验脚本随证据包归档。

3.8 第八组（campus10 · 端到端 · 优化）——评级 B

解读对象 `e2e_gpu/cpp/src` 引擎消费端链路，与优化角色对口。**亮点**：技术深度居八组之首：完整调用链总图（GUI/CLI 两条路径）、预处理 /前向/后处理逐行对账；耗时占比的“假设登记表”方法论（代码钉死事实 F1–F8 / 可复算算术 / 显式假设 A1–A3 三层拆分，结论以敏感性表而非单点成立）；**发现行级真实优化点**：`makeInput` 中 `cv::split` 写在通道循环内、每图冗余执行 3 次（流量约 330 MB 降至 180 MB，U3）；800 ms 预算书含“怎么测”与超支处置（U4）。**不足**：报告为 Cline 会话原文粘贴（含界面提示文字），无独立撰写的报告结构；人工核对痕迹散落（对/错/存疑标记不成表）；预算表前后两版数字不一致；组内仅三人一句话分工。**建议**：按全班五步模板重排报告主体，将预算书落为 `e2e_gpu/docs` 正式文档。

4 项目总评

实验一：八个组全部在截止前提交且对口完成，“AI 解读 + 人工核对 + 纠错”的入场达标目的已经达成——预期的项目理解**整体成立**，其中组 2/组 3/组 6/组 7 的产出（规格说明书、参数清单、口径修正、预算书）已经超出实验一本身，直接构成后续实验二/三的输入。

项目推进侧：截至本报告，八组在 4080 上尚未开展训练工作（等待任务统一下发，`incoming/` 机制已就绪）；本报告第 5 节的 U 清单与 DOC-04 的 M 清单合并执行后，即可形成 360code_cpp 的统一升级基线并向全员分发。

5 git 可更新点清单（来自八组工作，待确认后执行）

八组的实验一虽以“理解”为目的，但产出了若干**直接可入库**的修正与文档。汇总如下（与 DOC-04 的 M1–M5 相互独立，可并行确认执行）：

编号	来源	内容	建议动作	优先级
U1	组 3	`learned_params.h:194` 声明 gain 输入 16 维，而 `head_specs.py:28` 与 `onnx_provider.cpp:91` 均为 12 维（RGB 统计 [12]→ 增益 [6]）	以 12 为准修正 `learned_params.h`，消除 gain 头接线规格分歧	高
U2	组 6	`datasets/README.md` 批次登记缺 demo2（磁盘已存在）；README“已知平移真值”口径与代码事实（相位相关伪真值）不符	补登记 demo2 并标注合成批次；按代码事实修正标签口径描述	高
U3	组 8	`rop_stitch_engine.cpp:102-106` 的 `cv::split` 在通道循环内，每图冗余执行 3 次（预处理流量约 330 MB→180 MB）	将 split 提到循环外一次执行，跑基准 A/B 验证后入库	高
U4	组 8	`kPairBudgetMs`=800 ms/对与端到端总预算冲突；run 耗时只进判废不进日志	增加段级计时日志（对齐 [InitTiming] 风格）；预算常量讨论后调整	中
U5	组 7	`train.py:3` docstring“ResNet 风格”名不副实（无残差连接）；L32 转义序列触发 SyntaxWarning	改写 docstring（HomographyNet/UDIS2 风格）；修复转义	中
U6	组 1/2/5	理解性结论具备复用价值：p_correct 非校准概率、temporal 输入无 T 维、init_ms 不摊帧等	汇入 `docs/` 常见问题说明或对应 train.py 注释，防止后续组重复踩坑	中
U7	组 4	`label_builder.py` 三段 `label_source` 同名不同义、params 同名两用等字段口径风险	在 `label_builder.py` 顶部 docstring 写明字段口径表（或改名区分）	低
U8	组 3	增益补偿段 22 项写死参数清单 + 5 项设计存疑点（开链无闭环、限幅破坏对数域零和等）	作为 group07_gain 的设计输入归档至 `docs/`（或各组训练目录内的 DESIGN.md）	中

执行口径：U1–U3为代码/文档修正，改动小、收益明确，随 M1/M2 一并入库并在提交信息中注明来源组；U4/U6/U8 为增材（日志、文档、设计说明）；U5/U7 顺手修。清单将统一合并形成 360code_cpp 新基线，经总控台 `push` 分发全体 campus 账号，实现各组本地环境同步升级。

组	角色	评级	一句话点评
1	九头 • 匹配决策	B+	p_correct 纠错深刻，证据包与截图缺失
2	九头 • 特征质量	A+	标杆：识破提示词双陷阱，证据三件套最全
3	九头 • 缝合融合	A+	标杆：三通道对账 + 22 项参数清单 + U1 发现
4	数据与标签	A-	数据流图与字段口径扎实，规范细节有笔误
5	测试与评测	B+	指标方向表清晰，署名缺失、报告偏薄
6	端到端 • 数据	A	推翻任务单口径的伪真值发现 + 治理缺口
7	端到端 • 训练	A	闭环代数证伪 AI 误报，分工信息缺失
8	端到端 • 优化	B	深度最高（split 冗余/预算书），报告体例需重排

6 评级总表与执行说明

本报告为实验一验收点评的**正式通知**，自发布之日起生效，同步挂载于课程网站实验一提交界面。第 5 节 U1–U8 与 DOC-04 的 M1–M5 将按清单执行入库，合并形成 360code_cpp 统一升级基线后经总控台分发全体账号；各组点评中列出的整改事项（证据补齐、署名补全、报告重排等）请在实验二提交时一并完成。