

王一彤

☎电话: 15141251539

✉邮箱: 15141251539@163.com

🎂年龄: 23

📅毕业时间: 2026.07



🎓 教育经历

◇ **青岛理工大学** 建筑学 本科 2021.9-2026.7

💬 自我评价

- **多模态 AI 训练师全栈经验:** 头部 AI 公司 (智谱华章) 14 个月全职实习,覆盖 Agent 数据、视频 Caption、文生图评测、文生视频评测四大方向,深度参与 AutoGLM 沉思 / GLM-4.5V / CogView4 / GLM-Image / CogVideoX 等核心模型,熟悉 SFT 数据生产 + 模型评测全流程。
- **11 人评测团队管理经验:** 作为文生图评测组长统筹日常迭代评测与竞品对标,完整经历 GLM-Image(首个国产芯片训练 SOTA 多模态模型)发布评测,熟练运用双盲评测 / 量化评分 / Badcase 闭环等方法论。
- **半自动化流水线设计能力:** 基于"机器批量初判 + 人工聚焦边界样本"的 Human-in-the-loop 思路,在内部平台落地预标分流流水线,累计为团队减少重复劳动每周 40+ 小时;熟练使用 Codex / Claude Code / Cursor 等 AI 编程工具,了解 RAG / Transformer 架构等基础知识。
- **建筑学 + 绘画背景的视觉判断力:** 对构图、色彩、光影有专业敏感度,能在评测中快速定位画面问题。

📁 项目经历

◇ **CogVideoX 文生视频评测** 智谱 2026.3-2026.5

项目背景: 为企业级商用版 CogVideoX-2 大版本上线,完成评测集构建、多轮内部迭代评测及竞品对标(Vidu Q3 / 腾讯混元商用版 / MiniMax 海螺 2.3),辅助 PM 推进交付。

主要职责:

- 辅助 PM 制定评测方案: 围绕文本遵循 / 运动质量 / 画面质量三大维度,设计 GSB(3 档)+ 量化评分(1-5 分)+ 归因标签的混合评测法,制定跨模型对齐规则(统一 10s 档、静音、去标识盲评)
- 深度参与评测集构建: 真实用户 Query 打捞 + VBench + Prompt 工程三来源,覆盖六大商业场景、简中难 3: 4: 3 梯度,坚持"模型陌生、不含参考答案"出题原则
- 统筹 11 名评测员按周滚动交付: 抽检质检 + 争议复核, badcase 归因四大问题类型,周会同步算法并跟踪修复复评闭环,全周期零延期

项目成果: ▪ 小组累计完成评测约例 9000+ 条,产出约 18 份评测报告 ▪ 系统化整理 badcase 100+ 条,归因为 4 大核心问题类型并给出针对性优化建议,提交算法复查 ▪ 定位 CogVideoX 相对竞品核心差距(运动质量、画面清晰度、风格多样性等)并量化差距幅度,为下一轮迭代提供明确优化优先级排序

◇ **CogView4 与 GLM-Image 文生图评测** 智谱 2025.11-2026.2

项目背景: 智谱文生图架构转型期, CogView4 年度大版本 2025.12 发布, GLM-Image 2026.1.14 发布(自回归 + 扩散混合架构、首个国产芯片全程训练 SOTA 多模态模型)。承担日常迭代评测、中文文字专项、竞品对标三类工作。

主要职责:

- 统筹 11 人评测组(实习生 2 + 外包 8), 首周完成三套 500 题评测集构建与试评准入, 主导全流程: 需求承接 → 规则文档 → 评测方案 → 分差仲裁 → 报告产出 → 算法复查 → badcase 复投, 16 周零延期
- 设计 GSB(3 档)+ 量化评分(1-5 分)机制, 凭建筑/绘画背景承担审美类分歧仲裁;全周期产出约 20 份报告, 主笔正式

版 8 份(内部迭代 4 + 文字专项 2 + 竞品 2)

▪主导 GLM-Image 发布前 2 轮全量回归(P0 问题发布前全部召回)与发布后竞品对标(vs qwen-image-max / 千帆 ERNIE-iRAG), 验证新架构中文文字渲染代际优势

★搭建评测质效自动化工作流: 质效规则文档 → 脚本提示词 → AI 辅助编程产出数据脚本 → 输出人效/一致率/异常判定(CSV)→ 本地 Web 看板 → 自动推送飞书群, 支撑每日早会与成员定向校准

项目成果: ▪统筹完成评测用例 1 万+ 条、累计 2 万+(本人 1, 400+ 条, 以仲裁票与高难题为主) ▪全组一致率从约 82% 稳定至约 90%;badcase 闭环推动硬伤 badcase 率约 24%→15%;评测数据验证 GLM-Image 海报设计场景竞品对标第 1、进入国产文生图第一梯队

◇ 多模态视频 caption 质检 智谱

2025.7-2025.10

项目背景: 为 GLM-4.1V-Thinking / GLM-4.5V 视频理解训练产出数据: 10 秒短视频完成 6 维元标签(色调/角度/构图/光影/景别/镜头)+ 双层描述, 人工标注 20 分钟/条, 标注环节为产能瓶颈。

主要职责:

▪独立承担视频质检: 逐条填写质检结果(通过/修正/不通过/舍弃)与意见, 重点把控"描述事实性"(幻觉零容忍)与"时序锚点";维护 QA 文档、每日早会统一口径;新人前 3 批 100% 全检、熟练者降为 20%-30% 抽检

★搭建评测质效自动化工作流: 基于质检沉淀的 badcase, 主导设计"双 LLM 一致性分流 + 裁判模型兜底"预标流水线: 双视觉模型并行打标, 一致样本经裁判复核进轻校对、分歧样本进入人工精修, 经"离线验证→小流量→全量"上线

项目成果: ▪ 累计质检 2000+ 条, 质检准确率 95%+、描述事实性合格率 ≥98%;交付数据进入 GLM-4.5V 视频理解训练集▪ 流水线上线后标注人效 20 分钟/条 → 约 12 分钟/条(提升约 60%), 周产能约 450→750 条;两类高频错误初稿占比分别下降约 20% / 25%

◇ AutoGLM 多场景 Agent 数据标注 智谱

2025.3-2025.6

项目背景: 为 AutoGLM 沉思(2025.3.31 发布)工具调用、多步推理等 Agent 能力的 SFT 训练, 对多轮对话数据质检与改写, 覆盖旅游规划 / 餐饮预订 / 出行助手三大场景。

主要职责:

▪ 独立完成多轮对话(15 轮次/session)"思考 + 行动"双维度质检与改写;JSON 工具调用严格语法校验, 参数完整性 ≥98%;参与交叉质检(覆盖 ≥20%)

▪主导 3 名新人冷启动 SOP: 量化培训文档→疑难 case 入项试标考核→轮值在线答疑, 新人爬坡期从一周缩短至两天

项目成果: ▪ 16 批次零延期, 累计约 2 万+ 轮次(约 1, 400 session);整体准确率 95%+、JSON 校验通过率 99%;沉淀 QA 文档与 badcase 库持续复用, 为后续 GLM-4.5 Agent 能力训练做数据储备

在校经历

◇ 青岛理工大学 新媒体部 部长

2021.10-2023.10

▪ 统筹 27 人新媒体部多平台内容运营(公众号/视频号/抖音),策划视频大赛、AI 类校园宣传视频等活动;凭绘画与 PS 功底独立完成海报、封面与排版设计

▪ 运营学校抖音号:单条视频浏览量破 2W,任期内粉丝 2.1W→3.9W(增长约 85%)

相关技能

英语 CET-4; 计算机二级; 绘画; PS; CAD; 3D 建模; 团队管理

个人网站及作品集: <https://personal-portfolio-8zh.pages.dev/>