

Personal Resume

蔡中扬

联系电话：13840255823（微信同号）

邮箱：3968211268@qq.com

学历：辽宁传媒学院 | 视觉传达设计（本科）

身份：2027届在读 | 随时到岗

个站：https://www.czyxcn/



个人优势

AI 数据训练经验：1年多模态数据处理经验，参与百度图片Caption数据生产与生数科技多模态模型评测执行，覆盖需求拆解、试标准入、质检验收与交付复盘等环节。

PE 与自动化：精通Prompt Engineering、熟练运用Dify Workflow、CSV 批处理和主流 Vibe Coding 工具；可以从业务痛点和低效点切入，把重复性人工 SOP 快速转化为自动化工作流与提效代码脚本，显著降低人工重复工作量。

设计与艺术史：具备专业的审美判断、画面信息拆解与结构化表达能力；能从构图合理性、光影关系、透视合理性等维度判断图片和视频质量；并且利用对中西方艺术史的专业积累，搭建艺术风格图库和艺术专项问答 Agent，提升标注环节对各个艺术风格的识别效率与判别准确率。

Badcase 复盘：能基于自动化清洗/标注中的 Badcase 定位问题并修正脚本规则、系统提示词、temperature、Top-p 等参数，从而降低图片误杀率，提升预标注质量和机标占比；有意识地在收集评测过程中的 Badcase，提升评测集线上分布代表性和评测模型优化建议。

AI 行业敏感度：持续跟进多模态生成、模型评测、提示词工程等 AI 前沿动态，定期研读 Google DeepMind 技术成果与 arXiv 预印本；能够结合业务场景提出优化思路与创新方案，为数据生产与评测流程迭代提供参考建议。

实习经历

百度—图片 caption—项目助理

2025.07—2026.01

项目背景：项目处于文心大模型 5.0 上线前的多模态数据生产阶段，需要高质量图片 Caption 数据支撑模型在图文对齐、视觉内容理解和艺术风格识别等能力上的训练优化。针对原始图片质量差异大、人工描述效率低等问题，本人作为项目助理，搭建Dify清洗/标注wrokflow、动态准入集构建和艺术风格图片集，协助提升数据生产效率和交付质量。

核心职责：

- 图片清洗工作流搭建：**基于业务痛点初始图片数据质量差异大，容易造成数据分发不均，通过“脚本预筛 + VLM复查+灰度区人工审核”的Dify 自动化清洗工作流，流程对图片素材进行清洗从而让标注同学拿到稳定的可以标注的合格图片。全量上线后误杀率 5%。只需每日进行灰度区数据审核和根据 Badcase 迭代工作流。
- 图片 Caption 机标工作流：**针对人工直接描述图片时存在的人效低、词汇覆盖不足等问题，基于 Dify 搭建图片描述流程Workflow，通过工作流的不断迭代，辅助标注人员从“人工逐张描述”转为“模型预标注 + 人工校验改写”再到工作流进行数据分流，简单数据使用机标，标注人员只改写分流出的模糊 case、边缘 case和异常 case，实现机标率47%，改写部分人效累计提升 170%。
- 图片 Caption 标注：**参与多模态图片描述数据标注，围绕视觉主体、构图关系、光影色彩、艺术风格等维度输出结构化Caption；通过规则执行和 Badcase 归因，减少误标和主观化描述，标注准确率 94%。

- **艺术风格专项培训辅助：**针对艺术风格类 Caption 标注**主观性强、理解模糊**的问题，整理**印象派、波普艺术、浮世绘**等风格图片参考集；帮助标注人员快速建立风格判断标准，提升标注准确率；并且针对每周 Badcase 分析中所更新的规则文档，录制规则讲解视频并沉淀**培训视频库和 QA 答疑文档**，减少重复培训成本，专项标注准确率由 **85% 提升至 92%**
- **动态准入测试集构建：**新人入项后约有 4 天的准确率爬坡周期，发现原准入测试集与真实生产数据**样态偏差**；基于每日典型 Badcase 汇总，在一周内构建更贴近当前生产样态的**动态准入测试集**，并通过本地脚本按新增典型 case 数量等量替换旧 case，保证测试集**时效性**和与真实生产数据的**贴合性**。优化后，缩短新人入场爬坡周期3天，新人约在 1 天左右达到生产准确率。

项目成果：

- 图片 Caption 数据在上线前 SFT 阶段的清洗、标注、准入流程优化；
- 清洗工作流处理图片约 8 万张，识别不合格图片约 3 万张；其中脚本预筛拦截约 1.7 万张，脚本通过后由 VLM 拦截约 1.3 万张。数据清洗环节只需定时少样本抽检和专人处理每日灰度区数据；
- 图片预标注流程机标率达到 47%且预标注质量逐步提升，整体人效从17/h提升至 46/h。
- 艺术风格专项准确率由 85% 提升至 92%，新人准入爬坡周期由约 4 天缩短至约 1 天。

生数科技 多模态模型评测—执行岗

2026.01—2026.05

项目背景：项目处于 Vidu Q3 模型迭代期，针对视频生成、视频编辑和竞品横评等任务持续评测，关注指令遵循、光影协调度，叙事流畅性等核心能力。本人作为评测执行岗，参与版本评测、评测集维护等，为模型迭代、功能优化和评测维度完善提供评测侧依据。

核心职责：

- **多模态视频模型评测执行：**参与多个阶段性版本的视频生成模型评测，按既定评测标准完成**视频生成、视频编辑和竞品对比任务**；对模型输出结果进行质量判断、问题 Case 记录和结果整理，个人评测**准确率96%**。
- **视频生成专项评测执行：**围绕 T2V / I2V / R2V 等任务形态，按评测规则拆解不同任务的**重点关注项覆盖指令遵循、运动合理性和叙事流畅性等**；在同一 query 的竞品横评中，对比不同模型的效果差异，归纳优势场景、短板表现和可复查 Badcase。
- **评测集构建与样本维护：**参与评测集构建和动态维护，从用户 query、真实生产样本和 PE 自构建 Case 中整理样本来源；协助补充模型**陌生场景和不同难度梯度**样本，覆盖基础样本、中档样本和难例样本，提升评测集对线上分布代表性和模型短板场景的覆盖能力。
- **Badcase 问题归因：**对评测过程中发现的异常输出进行问题归因，围绕指令未遵循、参考一致性破坏、运动异常和画面稳定性问题沉淀典型 Case 与判定依据，提高后续复盘、问题回传和报告整理效率。
- **评测报告与迭代建议支持：**参与版本评测结果，输出结构化问题清单、典型 Case 示例和模型短板分析；基于Badcase 分布，辅助补充测试样本和完善评测维度，为后续模型迭代和评测集更新提供评测侧材料。

项目成果：

- 熟悉“评测集构建—版本评测执行—评测结果整理—评测报告输出”的基础流程，具备独立撰写小版本评测报告的能力；
 - 完成多个小版本视频生成模型评测、功能专项评测和竞品对比评测，个人评测准确率达 96%，为模型问题定位、版本迭代和评测集更新提供稳定的评测侧支持。
-

案例展示



图片清洗 workflow



图片Caption机标

选择 CSV 文件 导出处理后文件 清空结果

处理完成，可预览结果并导出生成 AI 报告。

判定规则 展开

处理结果预览

总记录数 100

高风险 22

中风险 19

正常 59

员号	姓名	所属组	任务类型	统计周期	完成任务数	有效工时(h)	平均产量	任务完成率(%)	一次合格率(%)	严重错误率(%)	综合得分	问题标签	风险等级	异常原因	管理建议
116	邓辉	A组	文本分类	2026-06	5100	165.0	30.9	100.0	82.5	3.8	90.9	速度-质量倒挂	高	合格率偏低；严重错误率超警戒线；速度-质量倒挂	重点复盘针对性培训；安排专项纠偏；对是否存在漏量行
117	潘龙	A组	文本分类	2026-06	5350	168.0	31.8	100.0	80.0	4.5	89.6	速度-质量倒挂	高	合格率偏低；严重错误率超警戒线；速度-质量倒挂	重点复盘针对性培训；安排专项纠偏；对是否存在漏量行
118	江莹	A组	文本分类	2026-06	4950	163.0	30.4	100.0	83.5	3.2	91.4	速度-质量倒挂	高	合格率偏低；严重错误率超警戒线；速度-质量倒挂	重点复盘针对性培训；安排专项纠偏；对是否存在漏量行
119	许阳	A组	文本分类	2026-06	2800	165.0	17.0	67.5	94.5	0.6	91.2	低效型	中	平均产量偏低；任务完成率不足	排查操作节奏辅导；核查任务卡点问题

标注质量数据分析工具