

言子奇

17680232110 | 2517468410@qq.com

个人网站: [yanziqi.pages.dev](#)

大模型评测实习 | AI 应用评测 | Agent 质量验证

大数据科学与技术本科在读，聚焦大模型与 AI 应用评测，能够设计并执行测试集、分析结果、追踪质量问题和维护回归流程；将“输出好不好、版本能不能发布”拆成案例、指标、失败等级和可追溯证据。



核心能力

场景评测与标准设计：从业务场景提炼评测维度、固定案例、Schema、约束、失败等级和通过标准，区分硬性 blocker、人工复核项与系统异常。

测试执行与结果分析：执行测试集并整理数据指标，分析 Bad Case、版本新增/修复问题和规则变化，使用黑盒测试、故障注入与回归对比跟踪质量。

自动化与问题闭环：使用 Python/脚本化流程、结构化报告和证据链复现问题，推动修复与复测，使结论能够回到版本、输入、规则和执行记录。

项目经历

AI Quality Gate | AI 输出评测与版本发布门禁

个人项目 | 2026.09

- 针对 AI 新版本平均表现变好、却可能引入严重错误的问题，建立 20 个固定合成案例、失败分级、版本对比、人工复核和发布门禁流程。
- 实现 Schema、约束、引用定位和回归检查，输出 PASS/BLOCKED/REVIEW_REQUIRED 与失败证据；98 项测试通过，候选版本通过数增加但 blocker 增加时仍阻断发布。

ToolGate | Agent 工具调用安全与可验证审计

个人项目 | 2026.09

- 针对 Agent 越权、审批参数篡改和执行回包丢失等风险，将工具调用拆成身份、策略、审批、执行、对账和审计阶段。
- 实现审批指纹、幂等、UNKNOWN 对账、故障恢复和脱敏证据包；历史全量验收覆盖 168/168 个测试函数、66/66 项验收 ID，干净种子演示 10/10 通过。

FitAgent | 多模态结果边界与用户修正

个人项目 | 2026.07—至今

- 针对视觉结果容易被误读为确定结论的问题，设计 observed / uncertain / not_observable 三态；关键点不足时返回 null，不伪造精确数值。
- 用户确认、否定或跳过以追加事件保留，方案通过 v1/v2/v3 Diff 解释变化；浏览器本地处理 MediaPipe 关键点，原图不上传，138/138 条自动化断言通过。

Cogniq | 自然语言数据分析质量控制

团队项目、联调与验收 | 2026.04—2026.07

- 面向自然语言生成 SQL 可能越权、危险或无法核验的问题，参与数据接入、前后端联调、分析任务和结果展示验收。
- 梳理“字段与权限约束—SQL Guard—只读执行—结果与证据绑定”链路；项目获中国大学生计算机设计大赛福建省一等奖、全国三等奖。

技能与成果

AI 评测：评测维度与测试集设计、测试执行、数据指标整理、Bad Case 分类、版本回归、失败归因、质量跟踪和评测报告。

工程与数据：Python、JavaScript、SQL、Vue 3、FastAPI/Fastify、Spring Boot、SQLite/MySQL、Docker Compose；HTTP/SSE、接口测试、故障注入和日志排查。

AI 工具协作：持续使用 Codex、Claude、Cursor、ChatGPT 辅助评测脚本、测试复现、结果分析和证据整理；不把固定案例结果表述为线上准确率。

证书与成果：中国大学生计算机设计大赛 A02 大数据应用组福建省一等奖、全国三等奖；大学英语四级 436 分。

教育经历

宁德师范学院 | 大数据科学与技术 | 本科在读

2023.09—2027.06