刘梦帆

AI 测试开发 / 测试效能工程师 · 南京 · 约 7 年工作经验

13949994838 | 651548728@qq.com | mengfanliu.ggff.net

个人优势

· 约 7 年软件测试与工具开发经验，覆盖医疗 AI、显示设备及分布式业务，能够独立开展测试方案设计、自动化开发与交付验证。

· 在团队落地测试 Skills、禅道 MCP 与人工审核流程；报告质控在确认范围、评审关键产物后将原预计 14 个工作日的一轮测试在约 5 个工作日内完成。

· 具备 ASR、医学影像及 RAG 评测实践，能够开发评测脚本、报告工具及设备测试软件；选屏工具投入使用后，选屏及报告工作由约一周缩短至约 3 个工作日。

核心技能

开发与自动化：Python、C#、Java；Playwright、browser-use、TestNG、OkHttp3、Allure；接口 / UI 自动化、测试工具与报告开发。

AI 评测与工作流：ASR / TTS、CER / WER、医学影像质控、RAG / OCR、推理性能；Skills / MCP、AI 自审修订与人工审核；Claude Code、Codex。

部署与测试基础：Linux、Docker、MySQL、内网模型推理与 RAGFlow；JMeter、Postman；串口 / 显示器协议、CA 系列色度计、DICOM。

工作经历

南京巨鲨显示科技有限公司

2022.11 - 至今

测试工程师 / 测试工具开发

· 负责报告质控、语音助手、影像质控与 RAG 知识库的需求分析、功能 / 性能测试及模型效果评测，建设可复用的数据集、脚本和报告流程。

· 建设测试工具与内部质量平台，推广 AI 辅助测试流程；牵头与组内 3 名成员开发显示器协议自动化工具，支持协议定义、批量发送、用例导入及自动判定。

· 承担医疗显示器嵌入式、串口 / 协议、亮度色度及显示性能测试，配合 IEC / DIN 相关规范验证；将 AutoView 测试 Pattern 软件用于产线质检。

· 部署内网测试平台、RAGFlow、Embedding / Reranker 与 GPU 推理服务，沉淀部署、更新和模型替换脚本。

· 完成医学教学竞赛系统的 WebSocket 内外网长连接、并发答题及稳定性测试，支撑大型赛事测试保障。

千寻位置网络有限公司

2019.10 - 2022.11

软件测试工程师

· 负责业务中台、多端应用、FindPixel 三维重建及高精度轨迹业务的功能、接口、性能与自动化测试；基于 Java、TestNG、OkHttp3、Allure 维护接口回归，编写消息模拟、数据构造及问题复现工具。

· 搭建 Linux、阿里云与 K8s 分布式测试环境，验证 RocketMQ / RabbitMQ / Kafka 消息链路及数据一致性；参与 FindPixel 云端机型对比与私有化机房交付验证。

· 结合 MySQL、Charles、Postman、JMeter 分析业务链路，支持共享单车运营监管、位置网格告警等政企项目的迭代与复杂问题复现。

项目经历 • AI 评测与测试效能

以下项目在巨鲨任职期间分阶段开展，包含持续维护与迭代。

巨鲨 AI 报告质控

2026.09 - 至今

AI 测试开发 / 测试方案统筹

面向医学报告的智能质控产品。在紧急交付阶段接手测试，负责范围确认、方案制定、评审组织、关键场景执行与多轮交付。

- 确认测试范围、制定方案并组织评审；将人工测试准则沉淀为 Skills，覆盖需求分析、可测试性、逻辑 / 物理用例及分轮报告。
- 在 Skills 中加入“生成 → 自审 → 问题修订 → 人工确认”流程，AI 按要求检查并修改产物，本人审核通过后推进下一阶段，负责复杂场景、风险识别与最终质量判断。
- 复用 UIAutoTest 辅助高优先级功能测试，通过自研禅道 MCP 管理缺陷与分轮数据；执行关键场景并开发推理性能批测和报告脚本，支持部署方案对比。

交付结果：在确认测试范围、关键产物经评审的前提下，将原预计 14 个工作日的一轮人工测试任务在约 5 个工作日内完成。

巨鲨 AI 语音助手

2026.08 - 至今

ASR 评测方案 / 数据构建与自动化

覆盖医生口述报告与会议场景，验证实时转写、医学术语纠错、锁定窗口输入与会议总结。

- 构建约 10 小时评测集，合成约占 90%，真实录音 1 个多小时；分组统计 CER（字错误率）和 WER（词错误率），按真实 60%、合成 40% 权重综合评估。
- 在 RTX 3090 单卡上测试转写延迟、处理速度、并发限制与资源占用；先做基准测试记录最快表现，再阶梯增加并发定位瓶颈。
- 独立制定方案，开发 TTS / ASR 批量推理、参考文本比对与报告脚本；复用 UIAutoTest 留存截图和执行轨迹，提交功能、性能及算法评测报告。
- 分别统计真实与合成语音的 CER / WER，观察到真实录音的两项错误率明显更高。60% / 40% 为本项目采用的综合评估权重。
- 调研 ASR / TTS 论文及开源实现，部署开源 TTS，结合文本合成、语音克隆和人工录音构建数据，协调 GPU 资源。

交付结果：基于本人同类新产品纯手工测试经验估算约 15 个工作日，借助自动化在约 5 个工作日内独立完成一轮测试；录音采集由他人协助。

禅道 AI 综合工具平台

2026.03 - 至今

独立设计与开发 / 测试效能

基于 Flask、Vue 3、MySQL 与 Neo4j 开发的内部质量平台，整合质量数据查询与需求知识图谱。

- 基于 Flask / Vue 3 / MySQL 开发检索、质量看板与导出功能；以 Neo4j 存储需求知识图谱，支持关系查询及手动编辑，平台已在团队使用。
- 针对模型直接输出 JSON 格式不稳定，改用 Function Calling 提供图谱增删改查工具；增加格式质量门禁与最多 3 次重试，改善图谱生成稳定性。
- 封装禅道 MCP，支持缺陷查询、批量新建、更新、关闭及质量数据读取，联动需求分析、用例编制与报告流程。

项目经历 • 工具开发与其他评测

选屏测试工具（ScreenSelect）

2025.06 - 至今

总体架构 / 核心后端独立开发

基于 C# / .NET 开发的医疗显示器面板筛选工具，连接设备测量、结果计算、数据保存与报告导出。

· 负责总体架构、核心后端与部分前端；基于 C# / .NET 接入厂商色度计 SDK，通过串口、HID、DDC 与显示器通信，处理断连、超时及重试，提供提示与重建入口。

· 实现亮度 / 色度均匀性、色坐标与色域计算，以及结果保存和多字段报告写入；借助 AI 辅助开发，导出 Excel、色域图及 Word 报告，重复采集覆盖上次结果。

交付结果：上线使用后，将原约一周的选屏及报告工作缩短至约 3 个工作日。

UIAutoTest 智能 UI 自动化框架

2026.03 - 至今

主导设计与 AI 辅助开发

基于 browser-use 主导 AI 辅助开发，增加执行编排，兼顾 AI 探索与脚本回归。

· 在 browser-use 外围增加执行编排、批量物理用例执行、自动重试、回放 JSON 与报告导出；外置模型接入和执行配置，以 Playwright 兜底，支持生成传统自动化脚本。已在语音助手和报告质控复用，关键步骤由人工复核。

巨鲨 AI 影像质控（含 CT 扩展）

2025.07 - 至今

模型评测 / 数据与评测自动化

覆盖常规影像及胸 / 头 / 腹部 CT 质控；每个质控项约 50 个样本。

· 学习标注规范，使用 Pari 及 AI 辅助标注，结合人工审核维护金标准。

· 用 Python 完成清洗、金标准匹配与评测，按质控项统计 TP / FP / FN / TN、准确率、召回率、F1 及误报漏报；生成 Excel / HTML 错误样本和指标报告。

· 使用 pynetdicom 构建 DICOM 推送与性能测试场景，支持不同模型版本的重复验证。

显示器性能与显卡综合测试工具

2023.06 - 至今

测试工具研发 / 视觉异常识别

训练使用正常画面；验证采用未参与训练的正常画面及由未训练数据构造的异常画面。

· 集成 FFmpeg 影像采集、日志与后处理；与 AI 协同完成数据筛选、清洗和方案设计，AI 辅助执行 EfficientNet-B0 / B3 迁移学习，结合图像规则筛选黑屏、白屏、花屏和 Logo 异常。

· 异常时自动暂停并保留上电与故障现场，人工确认后继续；自动汇总可疑片段与报告，减少逐段检查录像。

· 集成第三方 CLI，覆盖模拟插拔、显示器休眠 / 唤醒、屏保唤醒与主机重启，形成设备事件模拟和异常排查流程。

巨鲨 AI 知识库

2025.06 - 至今

RAG / OCR 质量评测

面向医疗文档处理、OCR、检索召回及知识问答的 RAG 应用。

· 以金标准书名比对检索结果计算召回率；按重排保留书籍中金标准相关书籍的占比评价重排序。

· 批量提问后人工评价回答；综合评分中回答占 50%，向量检索与重排序合计占 50%，配套数据清洗、批量评测与报告脚本。

教育经历

南开大学 · 计算机科学与技术 · 本科 2020 - 2022

济源职业技术学院 · 计算机科学与技术 · 大专 2016 - 2019