

李鹏斌

电话: +86 15673868311 | +081 070-8536-2519

邮箱: cralpbm@gmail.com

个人主页: <https://crabin.github.io/> 个人博客: <https://www.cnblogs.com/crabin>

Github主页: <https://github.com/crabin>



教育背景

岩手大学（日本国立大学）设计与媒体工学

2024.10 – 至今

- 硕士研究生 校级奖学金
- 平均绩点: 3.2/4
- 主修课程: 计算机视觉、融合设计、计算机网络系统、计算机动画、图像设计表现、图像合成。
- 研究方向: 基于生成对抗网络的入侵检测系统的研究。使用 Generative Adversarial Networks (生成对抗网络模型) 生成合成流量数据, 并与真实数据融合以进行 IDS 训练。旨在提升 IDS 的泛化能力, 并增强其对对抗性攻击的防御能力。

中南林业科技大学 信息与计算科学

2019.09 – 2023.06

- 学士
- 平均绩点: 3/5
- 主修课程: JAVA/C/C++/PYTHON语言程序设计、MySQL、计算机组成原理、操作系统、算法设计与分析、数据结构、信息安全原理及应用、大数据可视化技术、软件工程。

技术技能

开发技术

- 掌握 Python、Java、Go、C/C++ 等编程语言, 具备扎实的算法基础与良好的工程实现能力, 能够独立完成后端服务与 AI 应用开发
- 熟悉 Spring Boot、MySQL 等后端开发技术, 具备接口设计、数据处理、业务逻辑实现与系统联调经验, 参与过实际项目开发交付
- 熟悉 Linux、Git、Docker 等开发部署环境, 具备基础工程化开发、容器化部署与项目协作能力
- 具备较强的数据处理与特征工程能力, 熟悉特征选择、编码、归一化、类别不平衡处理及时序数据建模流程
- 熟悉 OpenCV 与计算机视觉基础技术, 能够完成图像预处理、语义分割、目标检测等任务开发
- 具备从需求分析、方案设计、实验实现到效果评估的完整项目实践能力, 能够面向真实场景持续优化系统性能

AI 应用开发

- 熟悉 PyTorch、Transformers、HuggingFace 等 AI 开发框架, 能够独立完成模型训练、微调、评估与实验分析
- 掌握 GAN、AutoEncoder、CNN、RNN、LSTM、GRU、ViT 等深度学习方法, 具备异常检测、入侵检测与视觉任务的研究和实践经验
- 熟悉 LoRA、RAG、Agent 等大模型应用开发技术, 能够构建知识增强、多工具调用与任务自动化的智能系统
- 具备 AI Agent 系统开发能力, 熟悉 Prompt Engineering、harness、loop、任务规划、记忆机制、工具调用及多智能体协作等关键能力模块
- 具备 LangChain / LangGraph Agent 开发经验, 了解 Deep Agents 智能体快速开发框架, 能够设计状态机编排、任务拆解、工具路由、反思补查和多轮对话链路。
- 能利用 Claude Code、Codex 等主流 AI 编程工具进行全流程开发
- 具备网络安全与人工智能交叉应用能力, 能够将生成模型、大语言模型与安全场景结合, 解决异常检测、漏洞分析与自动化辅助等问题
- 持续关注 LLM、Agent、多模态、生成式 AI 等前沿技术, 具备较强的快速学习能力与跨领域整合能力

工作经历

湖南科创信息技术股份有限公司 java 开发工程师

2023.07 – 2023.11

参与长沙市政府采购系统开发, 基于 Java、Spring Boot 和 MySQL 完成后台接口设计与数据处理, 实现标书多格式解析、关键信息抽取与检索功能; 同时参与需求分析、测试与问题修复。

项目经历

Trip Planner | AI 旅行规划系统

2026.06 – 2026.06

- 概述:** 基于 Vue 3 + TypeScript + FastAPI + LangGraph + LangChain + ChromaDB + Redis + SQLite 构建的中文 AI 旅行规划系统, 支持快速 itinerary 生成、深度旅行研究、浮动聊天助手、地图/天气 enrich、历史管理和 Markdown/PDF 导出, 形成从需求输入、研究生成、结构化展示到持续追问与自然语言改行程的完整闭环。
- 技术:** Vue 3、TypeScript、FastAPI、LangGraph、ChromaDB、Redis、SQLite、RAG、Harness

• Github 链接: https://github.com/crabin/Trip_Planner

• 亮点:

1. 设计 Destination Intelligence 深度规划 Agent, 将目的地、日期、预算、人数和偏好转化为后台异步研究任务, 构建“章节规划、联网检索、证据总结、反思补查、来源追踪”的多阶段工作流; 基于 8 份真实报告统计, 平均生成 5 个章节、约 103 条 sources、84 个 unique URLs, 平均耗时约 33 分钟, 保存产物完成率 8/8。
2. 实现 Report Itinerary Agent, 将深度规划 Markdown / 历史 Report 通过章节切块、LLM 结构化抽取、质量门控、缓存指纹和断点续跑转换为前端可消费 itinerary, 复用地图、天气、预算、导出和聊天编辑能力。
3. 构建浮动聊天机器人 Agent, 基于 LangGraph 设计 Ask / Search / Research / Update 多分支状态机, 支持 itinerary 问答、实时天气/景点/交通查询、多步骤研究、方案比较、风险核查和自然语言改行程。
4. 构建聊天长期记忆系统, 使用 SQLite 管理 semantic / episodic / perceptual 记忆, 支持 session / trip 作用域、显式记住/检索/删除/清空/统计/opt-out, 并在 Chroma 不可用时自动降级为关键词检索。
5. 优化 RAG 检索链路, 引入 query rewrite、目的地硬过滤和轻量 rerank; 15 条旅行场景离线消融中, Top1 命中率从 53.3% 提升至 93.3%, TopK 提升至 100%, 跨目的地污染片段从 17 条降为 0。
6. 接入 Redis 缓存 RAG 检索、高德地图 POI/地理编码/路线估算和天气查询等高频外部调用, 并设计 Redis 不可用时的优雅降级路径。
7. 建立 Agent 评测与回归体系, 参考 tau3-style harness 构建 YAML 对话 case、fixture / real 双模式和多维 grader; 最新真实全量评估在 real LLM + real tools 环境下覆盖 116 个 case、126 个 turn, case pass、turn pass、intent accuracy、tool action match 和 high-risk pass 均达到 100%。
8. 构建 SSE 流式响应评估脚本, 在模拟慢搜索场景下将首反馈平均延迟从 0.4145s 降至 0.0052s, 降低约 98.7%, 改善长耗时 Agent 任务等待体验。

企业级 Java AI Agent 平台二次开发: Tool Registry、Hybrid RAG 评测与安全增强

2026.06 – 2026.07

• **概述:** 基于 java-up-up/super-agent 持续优化, 保留原项目的企业级 AI Agent、RAG、文档治理与管理后台能力, 并针对检索闭环、可评测性、安全性、工具扩展性、运行可观测性和引用可追溯性补齐工程化能力。

• 技术: JAVA、RAG、Neo4j、redis、minio、kafka、pgvector、Elasticsearch

• Github 链接: <https://github.com/crabin/super-agent-java>

• 优化亮点:

1. 基于 Spring AI Alibaba ReactAgent 设计企业级 Tool Registry, 将硬编码 Function Call 改造为可插拔工具注册体系, 接入 MCP Adapter, 支持工具启停、超时、配额、审计与异常降级, 使新增工具无需修改 Agent 装配代码。
2. 构建 RAG 离线评测体系, 基于历史 Trace 与人工 golden set 计算 Recall@K、MRR、nDCG、引用命中率和耗时指标, 支持检索参数、rerank 和 prompt 版本 A/B 对比, 为 RAG 策略迭代提供回归基准。
3. 在 Hybrid RAG 链路中接入证据充分性判定, 设计最多一次的有界迭代检索机制, 实现“证据不足 -> 查询补全 -> 二次召回 -> 证据合并”的闭环, 并通过 Trace 与离线评测验证召回提升。
4. 为 RAG 与联网搜索链路增加 Prompt Injection 防护和引用一致性校验, 对外部证据进行不可执行隔离、风险规则检测和答案引用回溯, 降低模型受污染证据诱导的风险。

MyAgent | 基于 PI Agent 增强的防御性安全评估智能体

2026.05 – 2026.06

• **概述:** 基于 PI Agent 开源项目二次开发的 TypeScript CLI/TUI 智能体系统, 在保留原有代码阅读、编辑、命令执行、会话管理、多模型接入和扩展机制的基础上, 新增面向授权防御性安全评估的安全子代理能力。

• 技术: TypeScript、Node.js、PI Agent、CLI/TUI、Extension System、Prompt Engineering、Multi-Agent Workflow、Security Guardrails、Vitest

• Github 链接: <https://github.com/crabin/MyAgent>

• 亮点:

1. 接入 Security Subagent Extension, 在 PI Agent 会话启动链路中自动识别安全类请求, 并动态注入安全模式 Prompt、工具集和上下文记忆。
2. 构建授权范围控制机制, 要求用户明确目标资产、允许动作、授权目的和有效期限后, 才允许执行端口检查、安全头检查、网络发现等主动验证。
3. 设计多阶段安全工作流, 将任务拆分为资产发现、攻击面分析、PoC 研究、风险评估、修复建议和报告生成, 并通过 step / iteration plan 控制执行边界。
4. 扩展 Asset Discovery、Attack Surface Analysis、PoC Research 三个专用子代理, 分别负责资产基线、服务暴露面和漏洞资料归一化。
5. 增强安全工具体系, 支持安全记忆、被动情报检索、网页分析、受限爬取、CVE/NVD/Exploit-DB/MITRE 数据归一化和 Markdown / JSON 报告生成。
6. 使用 Vitest 测试覆盖意图识别、授权过期、工具注册、记忆持久化、多子代理委派、人工确认和报告生成等核心流程。

Paper Humanizer Skill | 中英文学术文本人性化润色工具

2026.01

- 概述：设计并开源面向中英文学术写作场景的 Claude Code Skill / Python CLI 工具，用于识别并弱化论文、技术报告中的 AI 写作痕迹，在严格保留事实、数据、引用和专业术语的前提下，提升文本自然度、学术表达质量与投稿可读性；项目在 GitHub 获得近 100 Stars（当前 96 Stars）。
- 技术：Claude Code Skill、Prompt Engineering
- Github 链接：<https://github.com/crabin/paper-humanizer-skill>
- 亮点：
 - 设计学术润色 Prompt 与用户模板，支持中英文识别、语气配置和四段式标准输出。
 - 构建 AI 痕迹短语黑名单，弱化模板化表达，提升文本自然度。
 - 加入严格事实保护约束，保留数值、结论、引用和专业术语，避免润色失真。
 - 提供 Python CLI 与示例数据，支持文本批处理和标准化 Prompt 生成

AutoPentest-XL 自主渗透测试多 Agent 平台

2026.01 – 2026.02

- 概述：设计并实现面向授权渗透测试场景的自主化安全测试平台，构建基于 PDEF（感知-决策-执行-反馈）闭环的多 Agent 协作框架，覆盖资产侦察、漏洞分析、利用尝试、权限提升与失败反思全流程；结合 RAG、图状态编排与安全围栏机制，提升自动化渗透测试的执行效率、路径规划能力与安全可控性。
- 技术：Python、LangGraph、LangChain、RAG、ChromaDB、Neo4j、Docker、Scapy
- 亮点：
 - 基于 LangGraph + PentestState 设计多 Agent 状态编排机制，实现 Commander / Scanner / Exploiter / Privesc / Reviewer 等角色解耦协作，并支持反射重试、压缩上下文和人工接管等动态路由。
 - 构建增强型 RAG 检索体系，使用 ChromaDB 存储工具规范与漏洞知识，结合 Neo4j 维护内网拓扑、攻击路径与长期记忆，支持 CVE 检索、GTF0Bins 映射及利用路径优化。
 - 实现安全围栏与沙箱执行机制，通过 CIDR 越界校验、危险命令拦截、Reviewer 审核和 Docker 隔离执行，确保所有测试行为限定在授权范围内，提升平台安全性与可审计性。
 - 引入幻觉率评估与监控告警，基于 Shadow Grader 对 Agent 输出进行忠实度校验，结合 OpenTelemetry、Prometheus 监控 Token 消耗、执行耗时与利用成功率，增强系统稳定性与可观测性。
- Github 链接：<https://github.com/crabin/AutoPentest-XL>

基于 Vision Transformer 的日本仙台市城市卫星图像语义分割/CV、语义分割

2024.08 – 2024.09

- 概述：使用 Python、PyTorch、ViT 及 OpenCV 对公开获取的仙台市 GPS 卫星图像进行处理，提出通过 LoRA 微调实现高分辨率语义分割，有效识别道路、植被、建筑和车辆等关键地物，成果可支持城市规划与灾后应急分析。
- 链接：<https://crabin.github.io/projects/2024-09-01-Satellite-Image-Semantic-Recognition>

个人陈述

我具备扎实的计算机基础和较强的工程实践能力，熟悉 Python、Java、Go 等多种编程语言，在机器学习与深度学习方向有较系统的学习和实践积累，能够将理论方法应用到实际项目中。

我目前主要关注 AI 应用与网络安全的结合方向，在入侵检测、数据增强以及深度学习模型应用方面有一定研究和实践经验。同时，我也持续关注大模型、RAG、AI Agent、Harness、Loop Engineering 等前沿方向，并尝试将这些技术用于实际系统开发。

我做事风格偏问题驱动，遇到复杂任务时习惯先拆解问题，再逐步推进实现和优化；学习新技术的速度较快，也愿意主动尝试新的开发范式。我比较喜欢 Vibe Coding，享受借助 AI 编程工具快速验证想法、迭代原型和提升效率的过程。

生活中我喜欢打篮球，这也让我更加重视团队协作、沟通配合和持续投入。对我来说，技术不仅是能力，也是长期热爱，希望未来能够在 AI 应用与安全相关方向不断积累，做出真正有价值的产品和系统。