

袁晨杰

北京/西雅图 | 邮箱: chenjy4@uw.edu | 186-1258-7299 | 个人主页 | 领英

聚焦 Agent Workflow、LLM Evaluation 与 AI Safety, 具备从 0→1 搭建 AI 应用与评测体系经验, 擅长将用户场景、模型能力边界与数据指标转化为可落地的产品方案。

教育经历

华盛顿大学, 西雅图, 华盛顿州, 美国 (University of Washington) (QS 前 100) 2023 年 9 月 - 2027 年 6 月
本科 | 专业: 人因设计与工程 (Human Centered Design & Engineering), 人机交互方向 荣誉: 校长奖学金, GPA: 3.89
相关课程: 机器学习 / 深度学习 / 数据科学 / 数据可视化 / 统计学 / 实验设计 / 人机交互 (Human-AI Interaction) / 数据分析

实习经验

AI 产品经理 | 字节跳动 北京 / 2026 年 6 月 - 至今

- 设计面向招聘需求澄清场景的 AI Agent 原型, 将模糊用人诉求转化为结构化需求初稿; 负责场景拆解、用户角色分工、Agent workflow、MVP 边界与评估指标设计, 覆盖模糊点识别、冲突检测、动态追问、能力映射与 HR 可交接性评估。

AI 产品经理 | 一堂教育, Y.AI 北京 (远程) / 2026 年 5 月 - 2026 年 6 月

- 参与建设服务 3 万 + 注册用户的企业内部推荐系统, 围绕“用户档案 + 业务标签 + 推荐策略”设计个性化内容分发方案。
- 协同业务与研发团队梳理推荐场景、用户画像和指标口径, 推动推荐能力在培训、内容运营和员工成长等内部业务场景中落地。

项目经历

AI 产品经理 | 字节跳动「AI 全栈开发挑战赛」RoundTable: 多 Agent 协作平台项目 北京 (远程) / 2026 年 5 月 - 2026 年 6 月

- 从 0 到 1 定义 RoundTable 产品方案, 聚焦 AI Coding Agent 的上下文断裂、任务分工不清与执行不可控问题, 设计「PM Orchestrator + Coding Agents + IM 工作区」核心链路。
- 设计 Workflow 编排、Agent 分工、Handoff Card、Artifact 管理与依赖追踪机制, 将 Agent 编程从单次黑盒调用转为可编排、可交接、可追踪的协作流程。
- 输出 PRD、Golden Path、功能规格与 ADR, 明确 Orchestrator / Adapter / Handoff / Artifact 模块边界, 并落地覆盖工作区、任务看板、状态追踪和产物预览的 MVP Demo。

AI 产品经理 | 字节跳动「飞书 AI 校园挑战赛」基于 IM 的协作 Agent 项目 北京 (远程) / 2026 年 4 月 - 2026 年 5 月

- 面向校园团队“需求不清、任务拆解低效、PPT 初稿产出慢”的痛点, 设计多 Agent 协作流程, 覆盖需求澄清、任务分解、内容生成、PPT 初稿与归因分析。
- 定义核心指标为“从需求输入到可提交方案初稿的完成时长”, 将平均交付时间从 4h 降至 23min (-90%)。
- 设计 Skill Router 与 Agent 编排机制, 支持根据任务类型动态调用不同 Agent, 并通过结构化上下文传递降低新增 Skill 接入成本。
- 设计 PPT 评审 Agent 与演示彩排 Agent, 自动输出结构化评审意见、表达反馈与修改建议, 支持方案评审前多轮迭代。

AI 产品研究助理 | 用户赋能实验室 (The User Empowerment Lab) 西雅图, 华盛顿州 / 2024 年 9 月 - 至今

核心项目 1: SusBench (Computer Use Agent 上线前安全评测体系) (2025 年 7 月 - 2026 年 2 月)

- 针对 Computer Use Agent 在真实网页任务中可能出现的误点击、任务偏移、被暗黑模式诱导等风险, 设计面向上线前验收的 Agent 安全评测框架。
- 构建覆盖 55+ 网站、313 条任务路径的真实网页环境基准数据集, 评估 Agent 在复杂 UI 中的任务完成率、风险触发率与决策一致性。
- 定义 9 类用户诱导风险 Dark Patterns, 用于识别 Agent 决策中的潜在误导与异常行为。
- 基于 5 个 SOTA 模型评测结果, 发现 CV-only 模型在复杂 UI 场景下显著落后于 CV+HTML 方案, 为多模态 Agent 产品设计与模型选型提供依据。
- 论文被 UIST 2026 Full Paper 接收 (二作)。

核心项目 2: SusFix (暗黑模式 UI 自动修复系统) (2026 年 3 月 - 至今)

- 设计基于多模态 LLM 的浏览器插件, 自动识别网页中的暗黑模式 UI, 并生成可解释的修复建议, 降低普通用户被误导操作的风险。
- 针对 LLM 直接修复带来的高成本与不稳定问题, 设计“精准命中 → 页面结构匹配 → LLM 兜底”的修复策略, 提升相似页面复用率。
- 定义误修率、覆盖率与用户可理解性为核心评估指标, 在修复效果与误伤风险之间建立验证机制。

技能

AI 产品能力: Agent Workflow、Prompt Engineering、LLM Evaluation、RAG Evaluation、LLM-as-Judge、Badcase Analysis、HITL

产品能力: 用户调研、PRD、Roadmap、竞品分析、A/B 测试、数据驱动迭代

数据与实验: 指标体系设计、漏斗分析、留存 / 转化分析、Python、Excel、Google Analytics

工具与技术: Figma、FigJam、Miro、Python、Pandas、NumPy、HTML / CSS / JS、LLM API