Stein



Agent 后训练 · Agentic Data · 扩散语言模型 · AI4S 多模态大模型

17280801850 · 25210980109@m.fudan.edu.cn · 上海 · 中共正式党员

01 教育经历

复旦大学（推免）｜大数据学院 2025.09 – 2028.06

人工智能方向 · 知识工场实验室 · 全日制硕士（在读）

课程：NLP 与大语言模型、多模态大模型算法、最优化理论、高维统计推断、深度学习与空间智能、AI 前沿算法。

杭州电子科技大学｜卓越学院 2021.09 – 2025.06

计算机科学与技术 · 理工实验班 · 全日制本科

GPA 4.59/5（专业前 1%）；国家奖学金、国家级大创负责人、数学建模国赛二等奖、杭电 ACM 集训队成员。

课程：高级数据结构与算法、计算机组成原理、数字电路、深度学习、操作系统、编译原理、工科数学分析。

02 科研论文

1. Semantic DLM+: Improving Diffusion Language Models through Bias-variance Trade-off in Transition Kernel Design.
   NeurIPS 2026 · 已录用 · 一作
2. On the Trainability of Masked Diffusion Language Models via Blockwise Locality.
   ACL 2027 · 在审 · 一作
3. ProtAgent: Step-Level Post-Training Agent for Protein Understanding.
   ARR 在审 · 一作
4. STM: A Spatio-Temporal Model for Dynamic Graph Fraud Detection.
   DASFAA（CCF-B）· 已录用 · 一作
5. CUA-Recovery: Benchmarking Error Recovery in Long-Horizon Computer-Use Agents.
   ICLR 2027 · 在审

03 实习经历

字节跳动｜Seed · AI Data 2026.09 – 至今

GUI Agent 数据算法实习生 · 数据合成与 RL 训练环境建设

- **专业软件训练环境构建与验证：**面向桌面 App 逆向重实现，搭建可复现的 Mock 环境，建立 Setup—GUI Rollout—Verify—Reset 验证链路，集成 Trace 回放与产物审计；通过正负例校验独立 Oracle，区分 App 缺陷、Agent 执行失败与环境异常，支撑 RL 任务准入、reward 判定及修复回归。
- **GUI 训练数据质检与 Badcase 归因：**面向 CAD、Blender、SketchUp 等专业软件任务，构建题目、附件、Checklist 与执行结果的自动校验及人工复核机制，实现 SFT/RL 差异化准入与批量交付。复核 **9,641 个 Trace 事件**，定位终态约束验收、GUI 状态感知与长程执行策略缺陷，分析无效重试及高耗时、高 Token 轨迹，为训练团队提供能力归因与定向数据补齐建议。
- **真实回流环境复建与 RL 任务转化：**结合 Badcase Trace、附件与已有产物，解析运行依赖及软件前置状态，构建冻结初始条件的环境包，在干净 VM 中重放关键动作与状态迁移；通过资产与动作双重校验验证复建保真，交付可追溯的环境与执行证据，支撑真实失败样本向 RL 训练任务转化。

阿里巴巴｜平台技术 · AIR 计划 2025.12 – 2026.09

Research Intern · AI4S 多模态大模型 / Agent

科学任务合成、Protein–Text 多模态模型与工具调用 Agent；项目与量化结果详见第二页。

04 科研与项目经历

AutoResearch ｜科学 Harbor 任务合成与后训练数据闭环

阿里巴巴 · 面向 Qwen 后训练的 Harbor 任务生产

- **闭环设计：**将专家 Skill / Protocol 转化为任务生成与评测规则，构建数据审计、任务生成、Calibration、Rollout 和 Evaluation 流水线，贯通合成、修订、SFT 与评测。
- **难度与质量：**设计递归任务修订流程，结合 Rollout 与 Rubric Judge 反馈调节难度；采用 Solver / Verifier 隔离、多路径校准与跨模型评分，识别评测缺陷、推理错误和 Reward Hacking。
- **规模化生产：**搭建支持并发调度、断点恢复与版本追踪的 Multi-Agent 批量生产系统，累计交付 **2,000+ 个 Harbor 任务包**。
- **训练验证：**合成数据经 Qwen3.8-27B SFT 验证后交付 Qwen 后训练团队，推动 Terminal-Bench 等指标提升；相关任务提交至 Terminal-Bench Sci 并通过一轮 review。

Holon ｜ Protein–Text 多模态统一大模型

阿里巴巴 · Qwen3 / MoT · 蛋白质理解与生成

- **多阶段训练：**参与蛋白预训练、多模态 CPT 与科学 SFT，支持 Protein ↔ Text 双向建模、T2P / P2T 生成及序列–结构联合学习，开展模型规模、预算和数据配比实验。
- **数据与优化：**构建 UniRef 序列与文本–蛋白配对数据的混合方案，开展 Token 配比、Muon 优化器及 Scaling Law 实验，完成 **0.6B 至 8B** 的规模扩展；通过 Fitness 与配对数据对照分析蛋白能力保持及条件建模收益。
- **指令微调：**参与 0.6B 模型六组 SFT 对照实验，蛋白质位点坐标判别准确率由 **15.3% 提升至 21.2%（+5.9 pp）**；通用与专业混合数据使 IFEval 较纯专业数据训练提升 **29.02 个百分点**。
- **统一评测：**建设 ProtText-Eval，覆盖文本引导的蛋白生成、编辑和理解，从指令对齐、序列 / 结构质量及下游效果评估模型，集成多类 Baseline，形成团队评测基线。

ProtAgent ｜蛋白质工具调用与步骤级强化学习

阿里巴巴 · Qwen3-8B ReAct Agent · ARR 在审

- **后训练流程：**基于 Qwen3-8B 构建面向五类蛋白质注释任务的 ReAct Agent，设计 Tool-Trajectory SFT + On-Policy RL 两阶段流程；SFT 学习工具选择、调用格式与答案生成，RL 通过真实工具交互优化模型的工具调用与回答行为。
- **Step-Level Credit Assignment：**设计 Batched Step-Level REINFORCE，分别为中间工具调用（重复调用惩罚、工具执行校验等）与最终答案设置 reward，按交互步骤计算 reward-to-go，并用于对应响应 token 的策略梯度更新，改善多步工具调用中的信用分配。
- **实验结果：**在五类任务均衡的 OOD-1000 上，成功率由 Base **14.1%**、SFT **40.2%** 提升至 **61.0%**，超过 DeepSeek-V3.2-Thinking 671B 的 41.6%，与 GPT-5.5 的 60.8% 相当。

05 其他实习经历

天翼交通科技 ｜ VLA 部门

2025.04 – 2025.07

智能驾驶算法实习生

- 设计并落地面向 4D 自动驾驶数据的自动预标注流水线，集成开源模型完成道路预测、3D 检测、语义生成与 Occupancy 预测，为大规模数据生产提供高质量标注。

06 专业技能

训练与推理	PyTorch、Transformers、Accelerate、DeepSpeed（ZeRO-2/3）、FSDP2；分布式训练与显存优化；vLLM、FlashAttention-2。
后训练与 Agent	SFT、DPO、PPO、GRPO、RLHF；奖励设计、偏好数据、LLM-as-a-Judge；ReAct 多轮工具调用与策略优化。
工程与平台	Docker、Linux、Git、W&B、Slurm、GPU 集群；训练脚本、实验追踪、分布式任务提交与问题排查。
编程与基础	C++、Python、MATLAB、Java；算法与数理基础扎实，杭电 ACM 集训队经历，本科核心数学与计算机课程成绩 95+。