Stein



Agent Post-Training • Agentic Data • Diffusion Language Models • Multimodal AI4S Foundation Models

17280801850 • 25210980109@m.fudan.edu.cn • Shanghai • Full CPC Member

01 Education

Fudan University (Recommended Admission) | School of Data Science 2025.09 – 2028.06

M.S. student in Artificial Intelligence • Knowledge Works Lab • Full-time

Coursework: NLP and LLMs, multimodal foundation model algorithms, optimization theory, high-dimensional statistical inference, deep learning and spatial intelligence, frontier AI algorithms.

Hangzhou Dianzi University | Honors College 2021.09 – 2025.06

B.Eng. in Computer Science and Technology • Science and Engineering Honors Program • Full-time

GPA 4.59/5 (top 1% in major); National Scholarship, national-level undergraduate innovation project lead, Second Prize in the Contemporary Undergraduate Mathematical Contest in Modeling (CUMCM), member of the HDU ACM training team.

Coursework: Advanced data structures and algorithms, computer organization, digital circuits, deep learning, operating systems, compiler principles, engineering mathematical analysis.

02 Publications

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

03 Experience

ByteDance | Seed • AI Data 2026.09 – Present

GUI Agent Data Algorithm Intern • Data synthesis and RL training environments

- **Training environments:** built reproducible mock desktop apps and the Setup–GUI Rollout–Verify–Reset chain with trace replay and artifact audits; tested independent oracles with positive and negative cases to distinguish app, agent and environment failures, supporting RL task admission, reward judgment and regression testing.
- **Data QC and failure analysis:** built automated checks and human review for CAD, Blender, SketchUp and other professional software tasks with distinct SFT/RL admission criteria; reviewed **9,641 trace events** to identify failures in final-state verification, GUI state perception and long-horizon execution, and recommended targeted training data.
- **Failure replay to RL tasks:** reconstructed environments from real failure traces, attachments and artifacts; froze initial states, replayed key actions in clean VMs and verified fidelity through asset/action checks, supporting the conversion of real failures into traceable RL training tasks.

Alibaba | Platform Technology • AIR Program 2025.12 – 2026.09

Research Intern • AI4S Multimodal Foundation Models / Agent

Scientific task synthesis, protein–text multimodal modeling and tool-using agents; projects and results are detailed below.

04 Research & Projects

AutoResearch | Scientific Harbor Task Synthesis and Post-Training Data Loop

Alibaba • Harbor task production for Qwen post-training

- **Closed loop:** converted expert skills/protocols into task-generation and evaluation rules; built data-audit, task-generation, calibration, rollout and evaluation pipelines linking synthesis, revision, SFT and benchmarking.
- **Quality control:** designed recursive task revision with rollout and Rubric Judge feedback; used solver/verifier separation, multi-path calibration and cross-model scoring to detect benchmark defects, reasoning errors and reward hacking.
- **Scaled production:** built a concurrent multi-agent batch system with checkpoint recovery and version tracking, delivering **2,000+ Harbor task packages**.
- **Training validation:** validated synthesized data through Qwen3.8-27B SFT before handoff to the Qwen post-training team, contributing to downstream metric improvements; related tasks were submitted to Terminal-Bench Sci and passed one review round.

Holon | Unified Protein–Text Multimodal Foundation Model

Alibaba • Qwen3 / MoT • Protein understanding and generation

- **Multi-stage training:** contributed to protein pretraining, multimodal CPT and scientific SFT for bidirectional Protein ↔ Text modeling, T2P/P2T generation and joint sequence–structure learning; ran scale, budget and data-mixture experiments.
- **Data and optimization:** mixed UniRef sequences with text–protein pairs; studied token ratios, the Muon optimizer and scaling laws, scaling models from **0.6B to 8B** and analyzing protein-capability retention under fitness/paired-data settings.
- **Instruction tuning:** contributed to six SFT comparison experiments on a 0.6B model; protein site-coordinate discrimination accuracy rose from **15.3% to 21.2% (+5.9 pp)**; mixed general/domain data improved IFEval by **29.02 percentage points** over domain-only training.
- **Unified evaluation:** built ProtText-Eval for text-guided protein generation, editing and understanding, covering instruction alignment, sequence/structure quality and downstream effects; integrated multiple baselines and established the team evaluation baseline.

ProtAgent | Protein Tool Use and Step-Level Reinforcement Learning

Alibaba • Qwen3-8B ReAct Agent • ARR under review

- **Post-training workflow:** built a Qwen3-8B ReAct agent for five protein annotation task families with Tool-Trajectory SFT + On-Policy RL; SFT trained tool selection, calling format and answer generation, while RL optimized tool use and answers through real tool interactions.
- **Step-level credit assignment:** designed Batched Step-Level REINFORCE with separate rewards for intermediate tool calls, including repeated-call penalties and tool-execution validation, and for final answers; computed reward-to-go by interaction step and applied it to policy-gradient updates over the corresponding response tokens.
- **Results:** on OOD-1000 balanced across five task families, success rate improved from Base **14.1%** and SFT **40.2%** to **61.0%**, surpassing DeepSeek-V3.2-Thinking 671B at 41.6% and performing comparably to GPT-5.5 at 60.8%.

05 Earlier Experience

Tianyi Transportation Technology | VLA Department

2025.04 – 2025.07

Autonomous Driving Algorithm Intern

- Designed and deployed an automatic pre-annotation pipeline for 4D autonomous-driving data, integrating open-source models for road prediction, 3D detection, semantic generation and occupancy prediction to provide high-quality annotations for large-scale data production.

06 Technical Skills

Training	PyTorch, Transformers, Accelerate, DeepSpeed (ZeRO-2/3), FSDP2; distributed training and memory optimization; vLLM and FlashAttention-2.
Agents	SFT, DPO, PPO, GRPO and RLHF; reward design, preference data construction, LLM-as-a-Judge evaluation, ReAct-style multi-turn tool use, rollout analysis and policy optimization.
Platforms	Docker, Linux, Git, W&B, Slurm and GPU clusters; training scripts, experiment tracking, distributed job submission, environment reproduction and troubleshooting.
Programming	C++, Python, MATLAB and Java; strong algorithms and mathematical foundations; HDU ACM training-team experience; 95+ scores in core undergraduate mathematics and computer-science courses.