YIXIONG CHEN

Mountain View, CA
ychen646@jh.edu Homepage Google Scholar GitHub

RESEARCH PROFILE

Researcher and engineer working across three connected areas of **long-horizon AI agents**: agent architectures and memory, agent policy optimization, and agent evaluation and environment design. Built an autonomous AI R&D agent at Google reaching **76.0% medal rate on the full MLE-bench suite**; author of **Agentic-DPO**, an offline policy optimization method matching online RL (GRPO) on τ-bench at **8.5× lower training cost**; first author of two agent benchmarks and a contributor to **Agents' Last Exam**. Published at ICLR, NeurIPS, CVPR, and MICCAI.

INDUSTRY RESEARCH EXPERIENCE

Google Mountain View, CA
Research Intern, AI Innovation & Research *May 2026 – Aug 2026*

- Built an **autonomous AI R&D agent** reaching **76.0% medal rate on the full MLE-bench suite**; the two-model Gemini system also matches a Claude Opus 4.7 + Claude Code baseline on PostTrainBench.
- Independently designed a science-to-engineering **multi-agent architecture**: evidence-grounded hypothesis formation, research planning, long-horizon implementation and review loops, experience memory, context compression, budget-aware experimentation, sentinel-based failure detection.
- Built the end-to-end execution and evaluation stack — sandboxed code execution, experiment orchestration, grader integration, trajectory analysis, automatic recovery from failed runs.

Waymo Mountain View, CA
Research Intern, AI Foundation Team *May 2025 – Sep. 2025*

- Developed an **LLM-based preference model** for autonomous-driving trajectories that converts scenes into structured natural-language evidence and scores safety, comfort, and efficiency
- Reached **72.6% agreement** with human raters on pairwise preference, shipping the annotated preference dataset and automated evaluation pipeline.

Google Mountain View, CA
Student Researcher, Health AI *Jun. 2024 – Nov. 2024*

- Developed **CoCa-CXR**, a longitudinal vision-language model using cross-image attention over current and prior examinations; curated a **1.5M-sample** image-report dataset and improved temporal disease-progression classification by **8.6%** over prior methods (MICCAI 2025).

PH.D. RESEARCH — JOHNS HOPKINS UNIVERSITY (AUG. 2023 – PRESENT)

Agent Policy Optimization and Post-Training
Agentic-DPO; Meissa *2025 – 2026*

- **Agentic-DPO**: offline policy optimization that converts expert trajectories into state-conditioned preference supervision, contrasting the expert action against the policy's own most-likely mistake at each state — **+25.3 / +22.2 / +18.0 pp** over SFT at 2B / 4B / 9B on τ-bench, StableToolBench, and Mind2Web.
- On τ-bench retail (9B), lifts success 21.7% → **41.4%**, **matching online GRPO (40.0%) at 8.5× lower per-step training cost** with no environment rollouts or reward model; reaches baseline performance on **25% of the expert data**. Open-sourced.
- **Meissa**: a 4B multimodal agent post-trained on **40K** distilled state-action-observation trajectories with three-tier stratified supervision over strategy selection and execution; matches or exceeds frontier agents in **10 of 16 settings across 13 benchmarks** at **25× fewer parameters** and **22× lower latency**. Open-sourced.

Long-Horizon Agent Memory and Failure Analysis
The Compliance Trap; MemTrapBench *2026*

- Diagnosed *how* web agents consume retrieved memory via an Entry–Propagation–Recovery framework, holding memory content fixed while varying delivery schedule across **5 models** on WebArena/BrowserGym. Agents comply with **conflicting memory at 63–72%** regardless of scale, costing up to **−25.5 pp** task success — **stronger agents lose more**.
- Built **MemTrapBench** (247 tasks, 5,928 trials) to isolate memory-consumption failures, with permutation tests and Holm–Bonferroni correction; derived a retry-on-fail memory policy recovering **+10.4 pp** on WebArena and closing **96–101%** of the schedule-oracle gap.

Agent Evaluation and Environment Design
DeepTumorVQA; Agents' Last Exam *2025 – 2026*

- **DeepTumorVQA**: a hierarchical tool-augmented agent benchmark (**476K questions**, 9,262 3D volumes, 42 subtypes) decomposing each task into four stages so failures attribute to a specific capability; built the ReAct tool environment with deterministic ground-truth traces. Open-sourced.
- Agent SFT on **20K tool trajectories** improved accuracy **+25.6 pp** over direct inference while cutting average steps 5.5 → 3.8 and step-budget exhaustion **23.2%** → **0.3%**; tool gains survive realistic (non-oracle) tools at **93% retention**.
- Contributed **15 verified healthcare-domain task instances** to **Agents' Last Exam** (1,490 long-horizon computer-use tasks, 88 institutions), adopted as a core evaluation in the GPT-5.6 and Kimi K3 releases.
- *Earlier work (2020–2023, JHU & SRIBD):* self-supervised learning, meta-learning, noisy-label learning, training dynamics — incl. *layer convergence bias* across MLPs, CNNs, and ViTs on eight datasets (**ICLR 2023**), plus CVPR, MICCAI, TMI, and ICASSP papers.

SELECTED PUBLICATIONS

1. **Chen Y**, et al. *Pro^2 Agent: Scientific Problem Probing for Autonomous AI R&D*. Manuscript in preparation, 2026.
2. **Chen Y**, Yuille A. *Agentic-DPO: From Imitation to Agentic Policy Optimization on Expert Trajectories*. **NeurIPS 2026**.
3. **Chen Y**, Bai X, Yuille A. *The Compliance Trap: Diagnosing How AI Agents Consume Conflicting Memory*. **NeurIPS 2026**.
4. **Chen Y**, Xiao W, Bassi PRAS, et al. *DeepTumorVQA: A Hierarchical 3D Benchmark for Tool-Augmented Agents*. arXiv.
5. **Chen Y**, Bai X, Pan Y, et al. *Meissa: Multi-modal Medical Agentic Intelligence*. Under review, **AAAI 2027**.
6. Sun Y, Han X, . . . , **Chen Y**, et al. (350+ authors). *Agents' Last Exam*. **NeurIPS 2026**. *(Healthcare task contributor.)*
7. **Chen Y**, Xu S, Sellergren A, et al. *CoCa-CXR: Contrastive Captioners Learn Temporal Structures for Chest X-ray VL Understanding*. **MICCAI 2025**.
8. Zhu J, **Chen Y**, et al. *MoLE: Human-centric Text-to-image Diffusion via Mixture of Low-rank Experts*. **NeurIPS 2024**.
9. **Chen Y**, Yuille A, Zhou Z. *Which Layer is Learning Faster? Layer-wise Convergence Rates in Deep Networks*. **ICLR 2023**.

Full publication list: Google Scholar.

EDUCATION

Johns Hopkins University — Ph.D., Computer Science; Advisor: Prof. Alan Yuille — Aug. 2023 – Dec. 2026
M.S.E., Computer Science; GPA 4.00/4.00 — May 2025
Fudan University — B.Sc., Data Science — Sep. 2016 – Jun. 2021

TECHNICAL SKILLS

Agents & Envs	Long-horizon tool use, multi-agent systems, ReAct, agent memory, context compression, sandboxed execution, benchmark/environment design; MLE-bench, τ-bench, WebArena, BrowserGym, Mind2Web, BFCL
Post-Training & RL	Offline policy optimization, DPO, GRPO, preference learning, reward modeling, trajectory distillation, SFT, LoRA, multimodal post-training
ML Systems	PyTorch, JAX, Hugging Face, TRL, DeepSpeed, Accelerate, vLLM, distributed training, SLURM; Python, Bash, Git, Linux, LaTeX