Email: yuexing@mit.edu | Website: https://yuexinghao.github.io/Yuexing-Hao/

YUEXING HAO

Summary

PostDoc in EECS at MIT, specializing in **multimodal, LLM personalization, post-training (RLHF, SFT), and agentic workflows**. Experienced in RAG pipelines, MCP agent development, and RL fine-tuning.

Core Technical Competencies

Multimodal AI & Foundation Models: Vision-language models (CLIP, Flamingo), text-to-image generation (Stable Diffusion), multi-modal RAG systems, cross-modal alignment, unified embedding spaces

LLM Alignment & Post-Training: RLHF (Reinforcement Learning from Human Feedback), DPO (Direct Preference Optimization), PPO, constitutional AI, safety evaluations, reward modeling, instruction tuning, SFT (Supervised Fine-Tuning)

Reinforcement Learning: Policy gradient methods, Q-learning, actor-critic architectures, reward shaping, multi-agent RL, online/offline RL, model-based RL, RL fine-tuning for LLMs

AI Agent Development: MCP (Model Context Protocol) agents, agentic workflows, tool use, ReAct/CoT reasoning, multi-turn dialogue systems, intent classification, context management

Software Engineering: Python, JavaScript/TypeScript, SQL, Docker, Kubernetes, AWS/GCP, Git, RESTful APIs, microservices, CI/CD, system design, performance optimization

Education

Postdoctoral Researcher. EECS, Massachusetts Institute of Technology — 01/2026 to Present
Laboratory for Information & Decision Systems (LIDS)
Advisor: Marzyeh Ghassemi

IvyPlus Exchange Ph.D Scholar. EECS, Massachusetts Institute of Technology — 09/2024 to 01/2026
Laboratory for Information & Decision Systems (LIDS)

Ph.D. Human-Centered Design, Cornell University — 09/2022 to 01/2026
Concentrations in AI, Human-AI Collaboration, and AI Alignment

M.S. Computer Science, Tufts University — 09/2020 to 01/2022

B.A. Computer Science, Rutgers University-New Brunswick — 09/2017 to 05/2020

Working History

Research Intern — 05/2025 to 12/2025

Google Research (Multimodal AI & Agentic Systems, host: Mike Schaekermann & Rory Sayres) – Mountain View, CA

I worked on a Gemini-based conversational agent, "*Wayfinder*" designed to support consumers in seeking health information more effectively.

- Engineered **Gemini-powered conversational health agent (Wayfinder, now in Gemini 3.1Pro)** with advanced intent understanding, multi-turn coherence tracking, and adaptive response generation using chain-of-thought reasoning and tool-augmented generation
- Built production **RAG pipeline** with MCP-based agents on Gemini CLI, processing 10k+ queries from OpenFDA dataset and improving structured reasoning accuracy from 54.7% → 83.2%
- Published **first-author CHI 2026 paper** on large-scale system evaluation and agentic workflow optimization for health information retrieval [publication]

Human Frontier Collective Specialist — 04/2025 to 09/2025

Scale AI – Remote

I worked on designing and evaluating complex, domain-specific benchmarks to rigorously assess AI model capabilities, with a focus on identifying limitations and improving performance.

- Architected **domain-specific evaluation benchmarks** for frontier model capabilities assessment, covering reasoning, instruction-following, safety, and multi-step task completion
- Built **RLHF experimental pipelines** with domain expert feedback loops, validating model behavior through preference learning and constitutional AI principles
- Improved inter-rater concordance from **44.3% → 72.4%** through systematic benchmark refinement, rubric clarification, and evaluator calibration protocols

AI Research Intern — 03/2024 to 01/2026

Mayo Clinic Radiation Oncology Department (host: Wei Liu) – Phoenix, AZ

My work focuses on advancing patient education through human-centered AI systems, "*MedEduChat*". [publication]

- Develop multi-expert evaluation frameworks and data instrumentation tools to **measure alignment, robustness, and multi-turn coherence**.
- Published **2 first-author papers in Nature Digital Medicine** on agent behavior analysis, comparative performance vs. clinical teams, and expert-in-the-loop frameworks for safe AI deployment

Selected Publications

1. **Y. Hao**, J. Holmes, M. Waddle, N. Yu, K. Vickers, H. Preston, D. Margolin, C. Loeckenhoff, A. Vashistha, M.Ghassemi, S. Kalantari, W. Liu. Personalizing Cancer Education for Patients Using an EHR-Integrated LLM Agent. *Nature Digital Medicine 2025 (**top-1 journal for LLM in health**).* *https://www.nature.com/articles/s41746-025-02166-0*

2. R. Sayres *, **Y. Hao** *, A. Ward, A. Wang, B. Freeman, S. Zhan, D. Ardila, J. Li, I.-C. Lee, A. Iurchenko, S. Kou, K. Badola, J. Hu, B. Kumar, K. Johnson, S. Vijay, J. Krogue, A. Hassidim, Y. Matias, D.R. Webster, S. Virmani, Y. Liu, Q. Duong, & M. Schaekermann. (2025). Towards Better Health Conversations: The Benefits of Context-Seeking. arXiv 2025 https://storage.googleapis.com/research-media/wayfinding-ai.pdf. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '26), 13–17 April, Barcelona, Spain (**top-1 conference for Human-Computer Interaction**).*

3. **Y. Hao**, Z. Qiu, J. Holmes, C.E. Löckenhoff, W. Liu, M. Ghassemi, S. Kalantari. Large Language Model Integrations in Cancer Decision-Making: A Systematic Review and Meta-Analysis. *Nature Digital Medicine 2025 (**top-1 journal for LLM in health**).* *https://www.nature.com/articles/s41746-025-01824-7*

4. R. Cai, Z. Liang, B. Xu, Z. Li, **Y. Hao**, Y. Chen. TAG: Type Auxiliary Guiding for Comment Generation. *58th Association for Computational Linguistics (ACL), 2020 (**top-1 conference for Natural Language Processing**).* *https://aclanthology.org/2020.acl-main.27.pdf*

5. **Y. Hao**, K. Alhamoud, H. Zhang, H. Jeong, G. Yan, I. Puri, P. Torr, M. Schaekermann, S. Kalantari, A.D. Stern, M. Ghassemi. MedPAIR: Measuring Physicians and AI Relevance Alignment in Medical Question Answering. *arXiv* 2025 https://www.arxiv.org/abs/2505.24040 (*Under Review in **top-1 journal** Nature Medicine*)

6. **Y. Hao** *, Y. Huang *, H. Zhang, C. Zhao, Z. Liang, P.P. Liang, L. Sun, Y. Zhao, S. Kalantari, X. Zhang, M. Ghassemi. The Role of Computing Resources in Publishing Foundation Model Research. *arXiv 2025* https://arxiv.org/abs/2510.13621 (*Under Review in **Nature/Science***)

Research Impact & Awards **(Link to All Awards)**

- **$263k+** in competitive research funding, including APF K. Anders Ericsson Dissertation Grant ($10k), PCCW Frank H.T. Rhodes Grant ($11k), OpenAI Researcher Access Program ($5k)
- 513 Google Scholar citations across 9 first-author publications in Nature Digital Medicine (2×), ACM CHI (3×), AAAI, CSCW (3×)
- Outstanding Reviewer & Associate Chair for ACM FAccT, CSCW, CHI (80+ papers reviewed)
- IEEE ComSoc Student Competition Second Prize | Interdisciplinary Contest in Modeling (ICM) Meritorious Winner

Publications & Services **(Google Scholar Citation: 513)**

Author of 9 first-author papers in leading venues, including Nature Digital Medicine (2X), ACM CHI (3X), AAAI (1X), and CSCW (3X). Outstanding Reviewer and Associate Chair for ACM FAccT, CSCW, and CHI, with over 80 papers reviewed.