Haoyang Liu

(+1) 312-292-7407 | hl57@illinois.edu
Liu-Hy | Google Scholar | haoyang-liu
Urbana, Illinois - 61801, United States

Education

- **University of Illinois at Urbana-Champaign** *Jun 2020 - Present*
  Ph.D. in Informatics Urbana, IL, USA
  - GPA: 3.71/4.00
  - Research Focus: Robust and data-efficient learning; LLM agents for scientific discovery, data curation, and human-agent collaboration.
- **Beijing University of Posts and Telecommunications** *Sep 2016 - Jun 2020*
  B.S. in Telecommunication Engineering Beijing, China
  - GPA: 3.75/4.00

Publications

1. **Liu, H.**, Li, Y., Wang, H. (2025). **GenoMAS: A Multi-Agent Framework for Scientific Discovery via Code-Driven Gene Expression Analysis**. Under review.
2. **Liu, H.**, Li, Y., Xing, T., Dalal, V., Li, L., He, J., Wang, H. (2025). **Dataset Distillation via the Wasserstein Metric**. In *the 2025 IEEE/CVF International Conference on Computer Vision (ICCV 2025)*.
3. **Liu, H.**, Chen, S., Zhang, Y., Wang, H. (2025). **GenoTEX: An LLM Agent Benchmark for Automated Genomic Data Analysis**. In *the 20th Machine Learning in Computational Biology Conference (MLCB 2025)*. (Oral 14.4%)
4. **Liu, H.**, Singh, A., Li, Y., Wang, H. (2025). **Approximate Nullspace Augmented Finetuning for Robust Vision Transformers**. In *Proceedings of the 2nd Conference on Parsimony and Learning (CPAL 2025)*. (Oral 12.7%)
5. Xue, E., Li, Y., **Liu, H.**, Shen, Y., Wang, H. (2025). **Towards Adversarially Robust Dataset Distillation by Curvature Regularization**. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence (AAAI 2025)*.
6. Zhang, P., **Liu, H.**, Li, C., Xie, X., Kim, S., Wang, H. (2024). **Foundation Model-oriented Robustness: Robust Image Model Evaluation with Pretrained Models**. In *Proceedings of the 12th International Conference on Learning Representations (ICLR 2024)*.
7. **Liu, H.**, Li, Y., Jian, J., Cheng, Y., Lu, J., Guo, S., Zhu, J., Zhang, M., Zhang, M., Wang, H. (2024). **Toward a Team of AI-made Scientists for Scientific Discovery from Gene Expression Data**. arXiv preprint.
8. **Liu, H.**, Chaudhary, M., Wang, H. (2023). **Towards Trustworthy and Aligned Machine Learning: A Data-centric Survey with Causality Perspectives**. arXiv preprint.
9. Zhang, T., **Liu, H.**, Zhang, P., Cheng, Y., Wang, H. (2023). **Beyond Pixels: Exploring Human-Readable SVG Generation for Simple Images with Vision Language Models**. arXiv preprint.
10. **Liu, H.**, Sarol, J., Kilicoglu, H. (2021). **UIUC_BioNLP at SemEval-2021 Task 11: A Cascade of Neural Models for Structuring Scholarly NLP Contributions**. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval 2021)*, pp. 1112-1121. (**Best System Paper Award**, 1/175)

Research Assistantship

- **University of Illinois at Urbana-Champaign** *January 2023 - June 2025*
  Research Assistant - Prof. Haohan Wang Urbana, IL, USA
  - **Dataset Distillation via Wasserstein Metric:** Proposed novel dataset distillation method drawing from optimal transport theory, achieving new state-of-the-art performance on benchmarks like ImageNet-1K
  - **AI-made Scientists for Scientific Discovery:** Developed multi-agent framework integrating LLMs for automated scientific discovery from genomic data, demonstrating novel gene expression pattern identification
  - **Robust Vision Transformers:** Proposed approximate nullspace finetuning method for vision transformers, improving adversarial, OOD robustness, and clean accuracy at the same time
  - **Foundation Model Robustness:** Co-authored ICLR 2024 paper on evaluating image model robustness dynamically with pretrained models, beyond static benchmark evaluation
- **University of Illinois at Urbana-Champaign** *September 2020 - May 2022*
  Research Assistant - Prof. Halil Kilicoglu Urbana, IL, USA
  - **SemEval-2021 NLP Contribution Graph:** Led winning team in extracting scholarly contributions from NLP papers, combining deep learning with rule-based methods (Best System Paper Award)
  - **Multi-task Citation Analysis:** Developed joint model for citation context extraction and sentiment classification with gradient-based parameter sharing mechanism for biomedical literature

Internship

- **Tsinghua National Laboratory for Information Science and Technology** *March 2019 - September 2019*
  Research Intern - Prof. Chunxiao Xing & Guigang Zhang Beijing, China
  - **Knee Osteoarthritis Knowledge Graph:** Collaboratively developed automated system to construct medical knowledge graphs from electronic health records, using CRF for entity recognition and RNN for relation extraction based on a combination of learned and hand-crafted features, achieving high accuracy on test set
  - **Deep Learning for Disease Prediction:** Collaboratively developed multimodal neural network combining 1D CNN for clinical notes with fully connected layers for tabular data to predict diabetes risk, implementing techniques to handle class imbalance

Skills

- **Programming Languages:** Python, C/C++, Java, R, MATLAB
- **Machine Learning Frameworks:** PyTorch, Hugging Face, Timm, Scikit-learn, Keras, TensorFlow
- **Graph Neural Networks:** PyTorch Geometric (PyG), Deep Graph Library (DGL)
- **Data Analysis & Databases:** Pandas, NumPy, SQL, Cypher (Neo4j)
- **Research Areas:** Dataset Distillation, LLM Agents, Trustworthy ML, Information Extraction
- **Research Skills:** Experimental Design, Scientific Writing, Reproducible Research, Peer Review

Honors and Awards

- **List of Teachers Ranked as Excellent** *Fall 2022*
  University of Illinois at Urbana-Champaign [🌐]
  - Recognized for outstanding teaching performance as Instructor of Record for IS 203
  - Selected based on student evaluations and teaching effectiveness
- **Best System Paper Award** *August 2021*
  SemEval 2021 - 15th International Workshop on Semantic Evaluation [🌐]
  - Awarded for exceptional system design in structuring scholarly NLP contributions
  - Competed against international research teams in computational linguistics

Presentations & Talks

- **UIUC_BioNLP System at SemEval-2021** *August 2021*
  15th International Workshop on Semantic Evaluation (Virtual)
  - Presented winning system for structuring scholarly NLP contributions
  - Demonstrated cascade neural model approach to international audience
- **Research Showcase - NLP in Scientific Literature** *November 2021*
  UIUC School of Information Sciences
  - Showcased research on automated scientific paper analysis
  - Engaged with faculty and students on future research directions

Professional Service & Teaching

Teaching Experience: Instructor of Record for IS 203: Analytical Foundations (2022-2023, *List of Teachers Ranked as Excellent*); Teaching Assistant for IS 597 TML: Trustworthy Machine Learning (2024), IS 507: Data, Statistical Models, and Information (2023)
Peer Review: Reviewer for ICML 2023, 2024, NeurIPS 2024, ICLR 2024-2026, CVPR 2025, 2026; Program Committee member for AAAI 2026, KDD 2026 D&B Track

Interests & Activities

Outside of research, I maintain an active lifestyle through swimming and practicing Tai Chi. I have a passion for music, particularly classical guitar and Hip Hop. I also have a keen interest in open-source software development.