Gabrielle Kaili-May Liu

kaili.liu@yale.edu | pybeebee.com | LinkedIn | GitHub | Google Scholar

Curriculum Vita

Education

Yale University, 2023–Present
Ph.D. in Computer Science
Advisor: Arman Cohan

Massachusetts Institute of Technology, 2019–2023
S.B. in Mathematics with Computer Science
S.B. in Brain and Cognitive Sciences
GPA: 5.0/5.0

Research Interests

LLM Uncertainty/Calibration, LLM Post-Training, LLM Evaluation, Synthetic Data Generation, Metacognition in LLMs

Availability

June 1, 2026–September 1, 2026

Professional Experience

Summer Camp for Applied Language Exploration (SCALE), Human Language Technology Center of Excellence (HLTCOE), Johns Hopkins University, Baltimore, MD, USA
Visiting Research Scholar, June 2025–August 2025

- Developed pipeline to generate challenging, uncheatable, multihop queries which are unanswerable over any given database, for use in adaptively benchmarking the refusal capabilities of leading retrieval augmented generation (RAG) systems.

Explorer of Perception and Attention (EPA) Laboratory, National Taiwan University, Taipei, ROC Taiwan
Visiting Researcher, June 2023–August 2023, June 2020–August 2021

- Established the first experimental procedure to reliably assess and quantify metacognition of emotion.

Collaborative, Learning, and Adaptive Robots (CLeAR) Lab, National University of Singapore, Singapore
Visiting Researcher, June 2022–August 2022

- Investigated trustworthy AI principles for agentic decision-making in evolving environments.

Publications

Liu, G., Li, B., Cohan, A., Walden, W., Yang, E. Evaluating Retrieval-Augmented Generation Systems on Unanswerable, Uncheatable, Realistic, Multi-hop Queries. Under Review at ECIR 2026. https://www.arxiv.org/abs/2510.11956

Dietz, L., Li, B., **Liu, G.,** Ju, J.-H., Yang, E. Lawrie, D., Walden, W., Mayfield, J. Incorporating Q&A Nuggets into Retrieval-Augmented Generation. Under Review at ECIR 2026.

Walden, W., Mason, M., Weller, O., Dietz, L., Conroy, J., Molino, N., Recknor, H., Li, B., **Liu, G.,** Hou, Y., Lawrie, D., Mayfield, J., Yang, E. Auto-ARGUE: LLM-Based Report Generation Evaluation. Under Review at ECIR 2026.
https://arxiv.org/abs/2509.26184

Liu, G., Yona, G., Caciularu, A., Szpektor, I., Rudner, T. G. J., Cohan, A. MetaFaith: Faithful Natural Language Uncertainty Expression in LLMs. EMNLP Main 2025.
https://aclanthology.org/2025.emnlp-main.1505

Bean, A. M., Kearns, R. O., Romanou, A., Hafner, F. S., Mayne, H., Batzner, J., Foroutan, N., Schmitz, C., Korgul, K., Batra, H., Deb, O., Beharry, E., Emde, C., Foster, T., Gausen, A., Grandury, M., Han, S., Hofmann, V., Ibrahim, L., Kim, H., Kirk, H. R., Lin, F., **Liu, G.,** Luettgau, L., Magomere, J., Rystrøm, J., Sotnikova, A., Yang, Y., Zhao, Y., Bibi, A., Bosselut, A., Clark, R., Cohan, A., Foerster, J. N., Gal, Y., Hale, S. A., Raji, I. D., Summerfield, C., Torr, P., Ududec, C., Rocher, L., Mahdi, A. Measuring what Matters: Construct Validity in Large Language Model Benchmarks. NeurIPS 2025.

https://arxiv.org/abs/2511.04703

Liu, G., Shi, B., Caciularu, A., Szpektor, I., Cohan, A. MDCure: A Scalable Pipeline for Multi-Document Instruction-Following. ACL Main 2025.
https://aclanthology.org/2025.acl-long.1418

Lee, H. H., **Liu, G.,** Chen, Y. C., Yeh, S. L. Exploring Quantitative Measures in Metacognition of Emotion. Scientific Reports 2024.
https://rdcu.be/dwIct

Liu, G. Perspectives on the Social Impacts of Reinforcement Learning with Human Feedback. 2023.
https://arxiv.org/abs/2303.02891

Liu, G. A Robot Rights Curriculum Informed by Western and Eastern Principles. MIT Social and Ethical Responsibilities of Computing Symposium 2023.
https://robotrights.webflow.io/

Schaeffer, R., **Liu, G.,** Du, Y., Linderman, S., & Fiete, I. R. Streaming Inference for Infinite Non-Stationary Clustering. ICLR Workshop on Agent Learning in Open-Endedness 2022 & Conference on Lifelong Learning Agents 2022.
https://arxiv.org/abs/2205.01212

Schaeffer, R., Du, Y., **Liu, G.,** & Fiete, I. Streaming Inference for Infinite Feature Models. ICML 2022.
https://proceedings.mlr.press/v162/schaeffer22a.html

Lee, H. H., **Liu, G.,** Yeh, S. L. I know I'm happy, and I'm right: Metacognition of emotion. European Conference on Visual Perception 2021.
https://journals.sagepub.com/toc/peca/50/1_suppl

Liu, G. Weight Friction: A Simple Method to Overcome Catastrophic Forgetting and Enable Continual Learning. 2019.
https://arxiv.org/abs/1908.01052

Liu, G. Evaluating Gammatone Frequency Cepstral Coefficients with Neural Networks for Emotion Recognition from Speech. 2018.
https://arxiv.org/abs/1806.09010

Academic Service

AAAI 2026, AAAI-AIA 2026, NeurIPS 2025, ICML 2025, ICLR 2025, AAAI 2025, LLMAgents @ ICLR 2024, ICML 2024, IJCAI 2024, JASA-EL 2023, NeurIPS 2023, MIT Committee on Curricula 2022-2023, NeurIPS 2021, NeurIPS 2020, ICML 2020

Teaching

Head Teaching Assistant, CPSC 477/577 Natural Language Processing, Spring 2025, Spring 2026
Department of Computer Science, Yale University, New Haven, CT

Instructor, Fundamentals of Group Theory, Spring 2021, Spring 2022, Spring 2023
MIT PRIMES Circle, Department of Mathematics, MIT, Cambridge, MA

Lab Assistant, 6.036 Introduction to Machine Learning, Spring 2021
Department of Electrical Engineering & Computer Science, MIT, Cambridge, MA

Teaching Assistant, Fundamentals of Scientific Writing, Summer 2019
Research Science Institute (RSI), MIT, Cambridge, MA

Grader, 18.02 Multivariable Calculus, Spring 2023
Department of Mathematics, MIT, Cambridge, MA

Honors & Awards

Graduate Research Fellowship, National Science Foundation 2023-2028
Teaching and Learning Award, MIT Department of Mathematics 2023
$5,000 Prize, Envisioning the Future of Computing, MIT Schwarzman College of Computing 2023
Fung Scholar, Fung Foundation 2022
Eloranta Fellow, MIT 2022
Women and Mathematics Program, Institute for Advanced Study 2022

SERC Scholar, MIT Schwarzman College of Computing 2021 & 2022
Undergraduate Academic Award, MIT Department of Brain and Cognitive Sciences 2022 & 2023
MIT International Science and Technology Initiatives Award 2020-2021, 2022, 2023
First Place, Systems Software, Intel International Science and Engineering Fair (ISEF) 2019
$25,000 Winner, Top 40 Finalist, Regeneron Science Talent Search (STS) 2019
$25,000 Winner, Third Place, Siemens Competition 2017
Research Science Institute Scholar (RSI) 2018
Community Innovation Award, Society for Science 2018
Intel Excellence in Computer Science Award 2016-2019

Outreach & Activities

Mentorship Program, Women and Gender Minorities in Science at Yale, 2023-2025
Facilitating the professional success of young women and gender minorities in STEM

Editor-in-Chief, MIT Undergraduate Research Journal, 2021 & 2022
MIT's only peer-reviewed scientific journal serving the undergraduate population

Mentorship Program Director, Women in EECS, 2021-2022
A community for women in EECS that supports, encourages, and empowers them to succeed

Resources Chair, MITxHarvard Women in AI, 2021-2022
A supportive community for cisgender women, transgender women, genderqueer individuals, and non-binary individuals pursuing studies or research in AI and machine learning at MIT and Harvard

Representative, Dean's Action Group on Social & Ethical Responsibilities of Computing, 2020
Action group to create pedagogical materials for use across all levels of instruction, in order facilitate the development of responsible "habits of mind and action" for those who create and deploy computing technologies, and the creation of technologies in the public interest

Vice President of Operations, Delta Phi Epsilon, 2021-2022
International sorority at MIT whose mission is to equip members with the leadership skills to create positive change in the community

EMT, MIT Emergency Medical Service, 2020-2021
Organization that provides exception emergency medical care and education to MIT and the surrounding community

Internships & Programs

Summer Camp for Applied Language Exploration (SCALE), HLTCOE, JHU, Summer 2025
MISTI-Taiwan, MIT, Summer 2023
Women and Mathematics Program, Institute for Advanced Study (IAS), Summer 2022
MISTI-Singapore, MIT, Summer 2022
MISTI-Taiwan, MIT, Summer 2020-Summer 2021
Canada/USA Mathcamp, Summer 2019
Research Science Institute, CEE, Summer 2018
Canada/USA Mathcamp, Summer 2017
TN Governor's School for Computational Physics, Summer 2017
MathILy, Summer 2016
MathPath, Summer 2015

Pre-Graduate Projects

Trustworthy Collaborative AI, Prof. Harold Soh, CLeAR Lab, Summer 2022
Department of Computer Science, National University of Singapore, Queenstown, Singapore

Computation and Learning with Brain Assemblies, Prof. Tomaso Poggio, Projects in the Science of Intelligence, Spring 2022
Department of Brain and Cognitive Sciences, MIT, Cambridge, MA

Efficient Streaming Inference for Infinite Nonparametric Models, Prof. Ila Fiete, Fiete Lab, Fall 2021-Spring 2022
Department of Brain and Cognitive Sciences, MIT, Cambridge, MA

Metacognition of Emotion, Prof. Su-Ling Yeh, EPA Lab, Summer 2020-Summer 2021
Department of Psychology, National Taiwan University, Taipei, Taiwan

Lifelong and Meta-Reinforcement Learning for Structured Action Spaces, Prof. Josh Tenenbaum, CoCoSci Lab, Fall 2019-Spring 2020
Department of Brain and Cognitive Sciences, MIT, Cambridge, MA

A Mathematical Framework for Learning Shared Representations for Transfer Learning, Prof. Lizhong Zheng, Research Science Institute, Summer 2018
Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA

Preventing Domestic Violence Using Emotion Recognition in Speech, 2018

Neural Networks without Multiplications, 2017

Anthropomorphic Facial Emotion Recognition and Generation Objective through Machine Learning, 2017

Recognizing Emotions Using a Physiologically Based Facial Landmark Detection Model and Machine Learning, 2016

Technical Skills

LLM Skills
Designing & analyzing LLM benchmarks, Synthetic data generation & filtering for SFT/RL, End-to-end post-training pipeline development (SFT, DPO/PPO/variants, RM), Designing & training reward models, Rubric-based reward modeling, Large-scale LLM inference, Data efficient training, Training data selection, Distributed LLM inference & training, Training hyperparameter optimization, Experiment tracking, Model versioning and artifact management, Checkpoint selection, LLM meta-evaluation, LLM uncertainty quantification, Designing human data annotation, LLM confidence calibration, LLM faithfulness, Analysis of LLM abstention and hallucination, LLM API integration, Retrieval Augmented Generation (RAG), RAG evaluation, LLM-as-a-Judge prompt & rubric design, Multi-document processing, Multi-document QA & summarization, Multi-hop reasoning, Multi-hop QA, Unanswerable QA evaluation, Scientific QA

Languages
Python, Julia, SQL, Linux, Java, MATLAB, R, Fortran, Rust, LaTeX

Platforms
Google Cloud Platform, AWS, HuggingFace, GitHub, GitLab, Google Colab, JupyterLab, slurm, docker, Jira, Confluence, OpenAI API, Together API, Cohere API, Google AI Studio, Gemini API, HuggingFace

Libraries
PyTorch, trainer, transformers, trl, accelerate, deepspeed, unsloth, bitsandbytes, peft, wandb, TensorFlow, TensorBoard, Keras, OpenCV, dlib, scikit-learn, scipy, numpy, matplotlib, pandas, seaborn, Pyro, multiprocess, LlamaIndex, ragas, LiteLLM, LangChain, vLLM