

Liqin Ye

Email: liqiny@gatech.edu | Webpage: liqinye.github.io | LinkedIn: [in](#) | GitHub: [G](#)

EDUCATION

- **Georgia Institute of Technology** Atlanta, GA, USA
Ph.D. in Machine Learning Aug 2024 - Jun 2029 (Expected)
M.S. in Computer Science Aug 2023 - Dec 2025
 - GPA: 3.85/4.00
 - Research Interests: LLM agents, LLM reasoning, LLMs in finance.
 - Advisors: Dr. Chao Zhang and Dr. Sudheer Chava
- **University of California, Irvine** Irvine, CA, USA
B.S. in Computer Science Aug 2019 - Jun 2023
 - GPA: 3.95/4.00 (Magna Cum Laude)
 - Core Courses: Deep Generative Models, Deep Learning, Graph Algorithms, Formal Languages & Automata
 - Advisor: Dr. Stephan Mandt

PROFESSIONAL EXPERIENCE

- **Rufus Post-training Team, Amazon** Palo Alto, CA, USA
Applied Scientist Intern, Mentor: Rongzhi Zhang, Manager: Qingyu Yin May 2026 - Present
Topic: Longitudinal Agentic RL Training for LLM Forecasters

RESEARCH PROJECTS

- **AlphaBet: Reinforcement Learning for LLM Forecasting Agents** May 2026 - Oct 2026
 - Developed an agentic RL framework that trains a single LLM policy to search, reason, and report probabilities across multiple pre-resolution dates using feedback from resolved outcomes.
 - Provided a theoretical analysis showing how GRPO reward standardization can distort probability targets under multi-class proper scoring rules, and empirically validated the findings through controlled ablations.
 - Established the properness of the agentic RL forecasting objective, showing that reward maximization minimizes the expected sum of evidence-acquisition and probability-reporting gaps.
 - Improved GLM-4.5-Air (106B-A12B) accuracy to **50.5%** on Forecast-Dojo and **57.9%** on FutureX, with **14.4%/13.7%** lower Brier scores, achieving performance competitive with Claude Sonnet 4.6 (high).
- **Forecast-Dojo: Replayable Environments for LLM Forecasting Agents** Feb 2026 - Jul 2026
 - Built a replayable environment for benchmarking and training LLM forecasting agents at fixed historical dates, with date-restricted SEARCH/READ/COMPUTE tools and optional belief memory.
 - Curated **1,568** Polymarket events with **6,122** forecast steps and **18.8M** dated news articles from CC-News, using temporal train/evaluation splits, evidence-sufficiency filtering, and content-level leakage audits.
 - Benchmarked **12** models across tool and memory settings; research tools lowered Brier score for all models, while memory reduced median API cost by **24%** across five proprietary models.
 - Applied supervised fine-tuning to Qwen3-30B-A3B-Thinking-2507 on selected forecasting trajectories, improving held-out accuracy from 34.8% to **42.8%** and Brier score from 0.924 to **0.749**.
- **Evolutionary Task Discovery for LLM Reasoning** Jun 2025 - Jan 2026
 - Formulated reasoning data generation as directed search over a manifold of algorithmic skill and complexity.
 - Designed two evolutionary operators, Attribute Mutation for complexity scaling and Skill Crossover for compositional logic discovery, with a dynamic Zone of Proximal Development filter to ensure learnability.
 - Achieved **+5.7%** on code and **+6.4%** on math benchmarks over Qwen3-4B Thinking, with up to **+8.2%** on competition-level AIME problems; consistent gains in both pass@1 and pass@8 confirm no diversity collapse.
 - Validated effectiveness and scalability across Qwen3-4B/8B base and Llama-3.2-3B/3.1-8B Instruct models.
- **Precise Attribute Intensity Control in LLMs** Nov 2024 - May 2025
 - Developed a test-time gradient-based representation editing method (PRE-CONTROL) that steers an LLM's hidden states toward multidimensional attribute intensity targets in closed form to meet diverse user preferences.
 - Trained a lightweight value function with TD- λ to predict final attribute scores from partial sequences.
 - Achieved **6.6–30.7%** exact target-hit success (up to **5.1 $\times$** the best baseline) on text and code generation tasks.
 - Improved Pareto frontier hypervolume from 7.54 to **12.66** while reducing computational overhead **8-fold**.
 - Designed a controllable distillation pipeline using PRE-CONTROL, reaching higher Pareto hypervolume (**16.81** vs. 15.27) with only **15k** vs. 50k samples and **2.1** vs. 7.8 GPU hours compared to Best-of- N distillation.

PUBLICATIONS

1. **Liqin Ye**, Rongzhi Zhang, Xin Liu, Zheng Li, Yuan He, Qingyu Yin, Priyanka Nigam, Chao Zhang. **AlphaBet: Reinforcement Learning for LLM Event Forecasting Agents**. Submitted to *ICLR* 2027.
2. **Liqin Ye***, Haorui Wang*, Fardin Ahmed, Rongzhi Zhang, ..., Chao Zhang. **Forecast-Dojo: Replayable Environments for Benchmarking and Training LLM Forecasting Agents**. Submitted to *ICLR* 2027.
3. **Liqin Ye**, Yanbin Yin, Michael Galarnyk, Yuzhao Heng, ..., Chao Zhang. **Evolutionary Task Discovery: Advancing Reasoning Frontiers via Skill Composition and Complexity Scaling**. Submitted to *NAACL* 2027.
4. Rongzhi Zhang*, **Liqin Ye***, Yuzhao Heng, Xiang Chen, Tong Yu, ..., Chao Zhang. **Precise Attribute Intensity Control in Large Language Models via Targeted Representation Editing**. Preprint.
5. **Liqin Ye**, Agam Shah, Chao Zhang, Sudheer Chava. **Calibrating Pre-trained Language Classifiers on LLM-generated Noisy Labels via Iterative Refinement**. In *Proceedings of the 31st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2025.
6. Xinyue Zhao, Ruiyi Zhang, **Liqin Ye**, Rui Cao, Pengtao Xie, Sudheer Chava. **Can LLMs Take the Pulse of the Economy? A Real-Time Evaluation of LLM Nowcasts on Macroeconomic Indicators**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Main Conference (EMNLP Oral)*, 2026.
7. Michael Galarnyk, ..., **Liqin Ye**, ..., Sudheer Chava. **IPO-Mine: A Toolkit and Dataset for Section-Structured Analysis of Long, Multimodal IPO Documents**. In *Proceedings of the 32nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) Datasets & Benchmarks Track*, 2026.
8. Agam Shah, **Liqin Ye**, ..., Wei Xu, Sudheer Chava. **Beyond the Reported Cutoff: Where Large Language Models Fall Short on Financial Knowledge**. In *Proceedings of the 2nd Conference on Language Modeling (COLM)*, 2025.
9. Agam Shah, ..., **Liqin Ye**, ..., Sudheer Chava. **Words That Unite The World: A Unified Framework for Deciphering Central Bank Communications Globally**. In *Proceedings of the 39th Annual Conference on Neural Information Processing Systems (NeurIPS) Datasets & Benchmarks Track*, 2025.

SKILLS

- **Programming Languages:** Python, C/C++, JavaScript, TypeScript, SQL, Scala, HTML
- **Frameworks & Libraries:** PyTorch, slime, verl, vLLM, SGLang, Ray, LlamaIndex, NumPy, Git, pandas
- **Languages:** Mandarin, English
- **Passions:** Snowboarding, Basketball, Calligraphy, Photography

SERVICES

- **Teaching:** CS 7646 Machine Learning for Trading, CS 7641 Machine Learning, MGT 8997 AI in Finance
- **Reviewer:** ICLR 2027, ARR 2027/2026/2024, KDD 2025, NeurIPS 2026/2024, Computational Economics
- **Conference:** AI and Future of Finance (Setup, Poster Session)

HONORS & AWARDS

- | | |
|---|------|
| • Workday AI PhD Fellowship , Workday AI Research | 2026 |
| • ICS Honor Student , University of California, Irvine | 2023 |