Naimeng Ye

609-510-5488
naimeng.ye@columbia.edu

Profile

I work on **scalable training and deployment of LLM agents**. I build synthetic tool ecosystems and verifiable post-training RL environments, and develop methods to reduce latency in multi-step agent workflows without sacrificing correctness. I also work on personalization, uncertainty for sequential models, and causal inference under network interference.

Education

Columbia University — NY, USA
Ph.D. in Operations Research — 2022–Present
- Advisors: Hongseok Namkoong, Tianyi Peng

Princeton University — NJ, USA
B.A. in Mathematics, High Honors; Phi Beta Kappa; Sigma Xi — 2018–2022
- Undergraduate thesis (Mengdi Wang): model-misspecified offline RL

Experience

Apple — CA, USA
Research Intern, Health AI/ML — May 2026–Aug. 2026

Selected Projects

Speculative Actions: A Lossless Framework for Faster Agentic Systems
Naimeng Ye*, Arnav Ahuja*, Georgios Liargkovas*, Yunan Lu*, Kostis Kaffes, Tianyi Peng
ICLR 2026 (**Oral**)
Built a framework that parallelizes multi-step agent workflows via speculative execution. Reduced end-to-end latency by up to 20% while preserving exact correctness via prediction-verification and caching.
Validated across diverse settings to demonstrate generality. Deployed at Cursor.

Exchangeable Sequence Models Can Naturally Quantify Uncertainty Over Latent Concepts
Naimeng Ye, Hongseok Namkoong
ICLR 2024 Workshop ME-FoMo; submitted to *Operations Research*
Showed exchangeable sequence models can naturally quantify uncertainty over latent structure with coherent probabilistic semantics.

Other Papers

1. **Differences-in-Neighbors for Network Interference in Experiments**
   Tianyi Peng, Naimeng Ye, Andrew Zheng (α-β order)
   ACM EC 2025; **Finalist, RMP Jeff McGill Student Paper Award 2025**
   Gave a practical causal estimator for treatment effects when units interfere through their network neighborhood.

2. **Seq@K: Evaluating Language Models Under Sequential Feedback**
   Thomson Yen*, Gilbert Yang*, Naimeng Ye*, Minghui Chen*, Dong Xia, Tianyi Shen, Yechan Jeon, Yaqin Chen, Daksh Mittal, Fangyu Wu, Hongseok Namkoong
   Under review

3. **SynthTools: A Framework for Scaling Synthetic Tools for Agent Development**
   Tommaso Castellani*, Naimeng Ye*, Daksh Mittal*, Thomson Yen*, Hongseok Namkoong
   Under review
   Built a scalable pipeline to generate and simulate 30–50k tool APIs and tasks for training and evaluating tool-use agents.

4. **PersonalLLM: Tailoring LLMs to Individual Preferences.**
   Tom Zollo*, Andrew Siah*, Naimeng Ye, Ang Li, Hongseok Namkoong
   ICLR 2025
   Adapted LLM behavior to individual users from interaction data without full retraining.

5. **AI Agents for Web Testing: A Case Study in the Wild.**
   Naimeng Ye*, Xiao Yu*, Ruize Xu*, Tianyi Peng, Zhou Yu
   NeurIPS 2025 LAW Workshop
   Demonstrated deploying LLM agents for real-world web testing.

6. **LatentGym: A Testbed For Cross-Task Experiential Learning With Controllable Latent Structure**
   Daksh Mittal*, Tommaso Castellani*, Thomson Yen*, Naimeng Ye, Fangyu Wu, Minghui Chen, Tiffany Cai, Emmanouil Koukoumidis, William Zeng, Hongseok Namkoong
   Under review

Awards and Honors

Deming Doctoral Fellowship (Columbia Business School, 2025–26) • Shapiro Prize, Pyka Physics Prize (Princeton) • CGMO Gold Medalist (2017)

Service

Reviewing: ICLR 2025–26; ICML 2026; NeurIPS 2024. Operations Reasearch.

Teaching: PhD Generative AI (Fall 2025); EMBA/MBA Managerial Statistics (2023–25); COS217 Intro. Programming Systems (Princeton, Spring 2020).