Sikai Cheng

📞 470-232-8313 ✉ sikaicheng@gatech.edu 💼 sikai-c 🔗 sikai-c.github.io

Education

Georgia Institute of Technology **Aug. 2024 – Present**
Ph.D. in Machine Learning, Advisor: Pascal Van Hentenryck *Atlanta, GA*

Georgia Institute of Technology **Aug. 2022 – Dec. 2023**
M.S. in Operations Research (GPA 3.91/4) *Atlanta, GA*

The Chinese University of Hong Kong **Aug. 2018 – May 2022**
B.S. in Statistics, Data Science (MGPA 3.92/4 Rank 1/199) *Shenzhen, China*

Skills

- **Research:** Time Series Forecasting, Large Language Models, Reinforcement Learning, Stochastic Optimization, Spatio-Temporal Pattern Mining
- **Skills:** Python (Pytorch/Lightning, scikit-learn, pandas, Keras), Cpp, CUDA(C), R, Matlab, Git, Latex, Bash, Linux, HuggingFace, HPC

Publications

1. **Cheng, S.**, Hijazi, A., Konak, J., Erera, A., & Van Hentenryck, P. (2025). SPOT: Spatio-Temporal Pattern Mining and Optimization for Load Consolidation in Freight Transportation Networks. *Proceedings of the IEEE International Conference on Data Mining (ICDM)*. [PDF]

2. Ye, T., **Cheng, S.**, Hijazi, A., & Van Hentenryck, P. (2025). Contextual stochastic optimization for omnichannel multi-courier order fulfillment under delivery time uncertainty. *Manufacturing & Service Operations Management (M&SOM)*. **Finalist, M&SOM Practice-based Research Competition, 2025**. [PDF]

3. Chen, H., Tao, S., Chen, J., Shen, W., Li, X., Yu, C., **Cheng, S.**, Zhu, X., & Li, X. (2023). Emergent collective intelligence from massive-agent cooperation and competition. *NeurIPS 2022 Deep RL Workshop*. [PDF] [Code]

Under Review

1. Klamkin, M., Deza, A., **Cheng, S.**, Zhao, H., & Van Hentenryck, P. (2025). DualSchool: How reliable are LLMs for optimization education? *Manuscript under review*. [PDF] [Code]

Experiences

AI Institute for Advances in Optimization, Georgia Institute of Technology **Aug. 2024 – Present**
Graduate Research Assistant *Atlanta, GA*

- Deep Learning for Time Series Forecasting — Project with a Leading Global Wireless Communications Company
  * Proposed lightweight yet expressive deep learning models for channel state information (CSI) prediction in 5G systems, demonstrating superior robustness and generalization across thousands of realistic scenarios while reducing computational cost by up to 5× compared to the SOTA LLM-based predictor.
- AI-Integrated Optimization — Projects with a Fortune 500 Logistics Company and a Top U.S. Retailer
  * Developed a spatio-temporal pattern mining and optimization integrated framework for freight load consolidation, reducing transportation costs and travel distance by ~50% on large-scale industrial datasets; framework deployed in real operations.
  * Developed a contextual stochastic optimization framework that integrates distributional delivery-time forecasts with robust order fulfillment models, providing actionable insights that improved service quality and customer satisfaction.
- Large Language Models
  * Investigated the reasoning and mathematical understanding limitations of large language models by co-developing a novel benchmarking dataset for dual generation in linear programming, revealing fundamental gaps in model reliability.

S.F. Express Co., Ltd. **May. 2023 - Aug. 2023**
Research Intern *Shenzhen, China*

- Developed a novel approach for transportation tasks by formulating them as a variant of the Capacitated Vehicle Routing Problem with time limitations (CVRP); implemented Tabu Search Heuristic and Gurobi-based MIP solver, improving profits by 10%.
- Integrated inventory and routing decisions into an Inventory Routing Problem (IRP); applied Lagrangian Relaxation Methods and Fix-Partition-Policy to optimize performance.
- Modeled transportation tasks as a graph-to-sequence problem; designed GCN encoder and pointer network decoder, achieving near-optimal solutions on real-world datasets.

Parametrix Technology Co., Ltd. **Dec. 2021 - May. 2022**

Research Intern *Shenzhen, China*

- Developed multi-agent reinforcement learning models for the Lux.AI Kaggle competition, enabling cooperation among agents through neighborhood-based and grid-based modeling.
- Implemented multi-stage training with knowledge distillation to leverage both dense and sparse reward signals.
- Built and deployed evaluation environments; final model achieved a win rate >90% against top-ranked Kaggle agents.
- Co-authored the paper *Emergent Collective Intelligence from Massive-Agent Cooperation and Competition*, later presented at NeurIPS 2022 Deep RL Workshop.

Awards

- Finalist, M&SOM Practice-based Research Competition, 2025
- Undergraduate Research Award, The Chinese University of Hong Kong, 2021 & 2022
- Academic Achievement Scholarship, The Chinese University of Hong Kong, 2020 & 2021