

JUNYANG CAI

caijunya@usc.edu ◊ (272) 788-0465 ◊ caidog1129.github.io ◊ Google Scholar

RESEARCH PROFILE

Ph.D. candidate at the University of Southern California specializing in machine learning for combinatorial optimization. My research spans generalizable learning-augmented MIP solvers, adaptive search, and LLM-guided algorithm design, with publications at IJCAI, CPAIOR (**Best Paper Award**), ECAI, and NeurIPS Workshop, with research experience at IBM Research and Fidelity's AI Center of Excellence and solver software released through the COIN-OR Foundation.

EDUCATION

University of Southern California

Los Angeles, CA

Ph.D. in Computer Science; M.S. in Computer Science

Aug 2023 – Dec 2027 (expected)

- Advisor: Bistra Dilkina
- Research Direction: ML for Optimization, Mixed-Integer Programming solver, and LLM-Guided Algorithm Design

Bucknell University

Lewisburg, PA

B.S. in Computer Science & Engineering; B.A. in Mathematics

Aug 2019 – May 2023

- Advisor: Thiago Serra
- GPA: 3.98 / 4.0, Summa Cum Laude

RESEARCH INTERNSHIP EXPERIENCE

Research Intern at IBM Research

Yorktown Heights, NY

AI Algorithms and Applications Group

May 2026 – Aug 2026

- Developed a hybrid quantum-classical framework that uses LLM-driven evolution to automatically discover classical search and reduction heuristics for quantum-assisted combinatorial optimization.
- Integrated the evolved heuristics into Quantum-Informed Recursive Optimization (QIRO) and Quantum Local Search (QLS), reducing runtime relative to the original heuristics across all three evaluated QOBLIB problem classes.

Research Intern at Fidelity Investments

Boston, MA

AI Center of Excellence

June 2025 – Aug 2025

- Created GALA, a global-constraint-aware agentic LLM framework that translates natural-language optimization problems into executable MiniZinc models.
- Designed seven specialized constraint agents and an assembler agent, evaluating constraint detection on all 567 instances from the Text2Zinc dataset.
- Evaluated the end-to-end framework on 110 verified problems across three LLMs, improving solve rate by up to 4.54 percentage points over chain-of-thought prompting.

Research Intern at Fidelity Investments

Boston, MA

AI Center of Excellence

June 2024 – Aug 2024

- Designed BALANS, an online adaptive meta-solver that uses multi-armed bandits to select among 16 large-neighborhood-search configurations without offline training.
- Improved 79 of 94 hard MIP benchmark instances, reducing SCIP's primal gap by more than 75% and primal integral by more than 50%.
- Developed a solver-agnostic architecture supporting SCIP and Gurobi, extended it to the parallel ParBalans framework, and released the software through the COIN-OR Foundation.

SELECTED PUBLICATIONS

1. Bingheng Li, **Junyang Cai**, Yupeng Zhang, Bistra Dilkina, Jayant Kalagnanam, Dzung T. Phan. FunL2O: LLM-Guided Feature Function Design for Learning to Optimize. *Under Review at AAAI*.

2. **Junyang Cai**, El Mehdi Er Raqabi, Pascal Van Hentenryck, Bistra Dilkina. ID-PaS+: Identity-Aware Predict-and-Search for General Mixed-Integer Linear Programs. *International Symposium on Combinatorial Search (SOCS)*, 2026.
3. **Junyang Cai**, Serdar Kadioglu, Bistra Dilkina. GALA: Global LLM Agents for Text-to-Model Translation. *NeurIPS 2025 MLxOR Workshop*.
4. **Junyang Cai**, Taoan Huang, Bistra Dilkina. Multi-task Representation Learning for Mixed-Integer Linear Programming. *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR)*, 2025. **Best Paper Award**.
5. **Junyang Cai**, Serdar Kadioglu, Bistra Dilkina. BALANS: Multi-Armed Bandits-based Adaptive Large Neighborhood Search for Mixed-Integer Programming Problems. *International Joint Conference on Artificial Intelligence (IJCAI)*, 2025.
6. **Junyang Cai**, Taoan Huang, Bistra Dilkina. Learning Backdoors for Mixed-Integer Programs with Contrastive Learning. *European Conference on Artificial Intelligence (ECAI)*, 2024.

RESEARCH EXPERIENCE

LLM-Guided Program Evolution for Learning to Optimize

May 2026 – Present

Collaboration between the University of Southern California and IBM Research

- Co-developed FunL2O, the first unified framework for automatically designing executable feature functions through LLM-driven program evolution across eight learning-to-optimize pipelines and four frontier LLMs.
- Led the MILP component by integrating LLM-evolved feature programs with Predict-and-Search and GNN-guided learned backdoor branching across three combinatorial optimization domains.
- Reduced Predict-and-Search primal integral by up to 73% and learned-backdoor wall time by up to 40% relative to handcrafted representations.

Generalizable Machine Learning for Mixed-Integer Programming

Aug 2023 – Present

Supervised by Bistra Dilkina, University of Southern California

- Developed graph attention and contrastive learning methods for shared MIP representations across three solver-guidance tasks and three combinatorial optimization domains.
- Proposed the first multi-task representation-learning framework for ML-guided MIP solving, reducing primal integral by 49–93% relative to Gurobi when generalizing to larger unseen instances.
- Currently developing a foundation model for MIP by pretraining shared representations across diverse problem distributions and adapting them to new domains and solver tasks with limited fine-tuning.

SERVICE AND TEACHING

Reviewer: NeurIPS (2026), AAAI (2026,2027), IAAI (2026, 2027), TPAMI, AI Magazine

Teaching Assistant: CSCI 618 Machine Learning for Combinatorial Optimization, USC (Spring 2026), CSCI 570 Analysis of Algorithms, USC (Fall 2026)

AWARDS AND FELLOWSHIPS

2025 CPAIOR 2025 Best Paper Award

2025 3rd AI4OPT Student Workshop 3-Minute Research Pitch Competition, First Place

2023 Viterbi School of Engineering / Graduate School Fellowship

2023 INFORMS Undergraduate Operations Research Prize, Finalist

TECHNICAL SKILLS

Languages: Python, C++, Java, C, MATLAB

Machine Learning: PyTorch, PyTorch Geometric, Scikit-learn, NumPy, SciPy

Optimization and Quantum: Gurobi, SCIP, CPLEX, MiniZinc, PySCIPOpt, gurobipy, Qiskit

LLM: Hugging Face Transformers, vLLM, OpenAI / Anthropic APIs, RAG, LoRA

Systems and Tools: Git, Linux, Docker, Slurm, Bash