Soroush Vahidi

Ph.D. Candidate in Computer Science, New Jersey Institute of Technology (NJIT) | Expected Aug. 2027

sv96@njit.edu | Website | LinkedIn | GitHub | Google Scholar | Hugging Face

PROFESSIONAL SUMMARY

Ph.D. candidate working at the intersection of AI systems, algorithms, and optimization. Research spans LLM serving and inference, learning-augmented algorithms, retrieval-assisted optimization, and graph/ranking methods. Publications include SIGKDD, ESA, Algorithmica, Ad Hoc Networks, IEEE HPEC, and an accepted sole-authored Journal of Supercomputing paper. Experienced in building reproducible research systems in Python/C++ with HPC, ML, and mathematical-optimization tooling.

EDUCATION

New Jersey Institute of Technology (NJIT) Jan. 2022 – Expected Aug. 2027
Ph.D. in Computer Science

Isfahan University of Technology Jun. 2015 – Jun. 2020
B.Sc. in Computer Engineering (Computer Architecture)

EXPERIENCE

Software Developer Intern — Canfield Scientific, Parsippany, NJ Jun. 2025 – Aug. 2025

- Implemented C++/Qt features for a commercial dermatological/medical-imaging platform, supporting image acquisition, annotation, lesion tracking, and reporting.
- Debugged and maintained real-time multi-camera components and image-processing pipelines under production constraints, collaborating closely with engineering and QA teams.

Graduate Research Assistant — New Jersey Institute of Technology Jan. 2022 – Mar. 2023

- Conducted research in combinatorial optimization, scheduling, graph algorithms, operations-research modeling, approximation/numerical algorithms, and parallel/high-performance computing.
- Formulated and analyzed approximation bounds and exact mathematical programs for complex network routing and resource-allocation problems.

Graduate Teaching Assistant — New Jersey Institute of Technology 2023 – Present

Courses: CS 610 Data Structures and Algorithms (multiple semesters, 2023–2026); DS 677 Deep Learning (Fall 2025); CS 388 Mobile Application Development / Android (Spring 2026).

- Graded programming assignments, conducted Android development lab reviews, evaluated student code submissions on GitHub, supported course delivery, and contributed to departmental automation tools.

PEER-REVIEWED PUBLICATIONS

- **S. Vahidi**. "Training-Free Ranking from Pairwise Comparisons via Acyclic Graph Construction." *The Journal of Supercomputing*, accepted, in production, 2026. (Sole author)
- B. Schieber, B. Samineni, **S. Vahidi**. "Interweaving Real-Time Jobs with Energy Harvesting to Maximize Throughput." *Algorithmica*, 2025.
- M. M. Islam, **S. Vahidi**, B. Schieber, S. B. Roy. "Promoting Fairness and Priority in Selecting k-Winners Using IRV." *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024.
- B. Schieber, **S. Vahidi**. "Approximating Connected Maximum Cuts via Local Search." *Proceedings of the Annual European Symposium on Algorithms (ESA)*, 2023.
- **S. Vahidi**, B. Schieber, Z. Du, D. A. Bader. "Parallel Longest Common SubSequence Analysis in Chapel." *Proceedings of the IEEE High Performance Extreme Computing Conference (HPEC)*, 2023.
- H. Falsafain, M. R. Heidarpour, **S. Vahidi**. "A Branch-and-Price Approach to a Variant of the Cognitive Radio Resource Allocation Problem." *Ad Hoc Networks*, vol. 132, 2022.

CURRENT MANUSCRIPTS

- **S. Vahidi**. "Decision-Aligned Eviction-Value Prediction for Learning-Augmented Caching." *Knowledge-Based Systems*, second revision (R2 under review).
- **S. Vahidi**. "Retrieval-Assisted Instantiation of Natural-Language Optimization Problems." Submitted to *SN Computer Science*.
- **S. Vahidi**. "Nominal Budgets and Realized Resources in Closed-API Large-Language-Model Inference: A Performance Evaluation Protocol." Submitted to *Performance Evaluation*.

SELECTED RESEARCH PROJECTS

LLM Serving / Scheduler Research Baselines | Selector Regret

- Evaluated scheduler interventions using trace-driven counterfactual replay and instrumented vLLM experiments, building reproducible analysis pipelines with selector-regret matrices and large-scale SLURM execution.

Retrieval-Assisted Optimization Modeling GitHub

- Designed retrieval-conditioned schema selection and deterministic grounding for mapping natural-language optimization problems to structured LP/ILP formulations, evaluated on a 331-query corpus.

Learning-Augmented Cache Eviction GitHub

- Developed decision-aligned eviction-value prediction models for online paging, evaluating performance across cache policies, prediction horizons, and workload shifts via trace-driven SLURM pipelines.

Training-Free Pairwise Ranking / Feedback Arc Set Heuristic Metrics | FAS Results

- Implemented training-free ranking algorithms via weighted feedback arc set and acyclic graph construction, releasing public metrics-only datasets for classical, GNNRank, and FAS-style heuristic comparisons.

Budgeted / Adaptive LLM Inference Allocation Metrics

- Analyzed cost-latency-accuracy tradeoffs under fixed logical-call budgets across closed-API providers (OpenAI, Gemini), releasing a 12,300-row metrics dataset on reasoning benchmarks (GSM8K, GPQA).

TECHNICAL SKILLS

Programming & Research Engineering: Python, C/C++, SQL, MATLAB, Chapel, Git/GitHub, Linux, pytest, CI/CD
AI/ML Systems: PyTorch, TensorFlow, scikit-learn, Hugging Face, vLLM, NumPy, pandas
Optimization & HPC: Gurobi, CPLEX, MiniZinc, SciPy/HiGHS, SLURM, OpenMP, CUDA, graph algorithms, parallel computing

ACADEMIC SERVICE

Reviewer, NeurIPS 2026; Program Committee Member/Reviewer, ACM ICMR 2026; extensive journal peer-review service across AI/ML, algorithms, optimization, operations research, and theoretical computer science.

HONORS & AWARDS

- **Best Paper Award**, WALCOM 2023, for "Interweaving Real-Time Jobs with Energy Harvesting to Maximize Throughput".
- **ACM-ICPC Asia Tehran Regional** — 17th Place, 2017; Honorable Mention, 2015–2016; Regional Competitor, 2015–2017.