

Zohair Shafi

✉ shafi.z [at] northeastern [dot] edu | 🌐 zohairshafi.github.io | 💬 zohairshafi | 💬 zohairshafi | 🌐 Boston, Massachusetts, U.S.

Education

Northeastern University

PhD in Computer Sciences (Machine Learning and Network Science) - GPA 3.98/4

Boston, MA, U.S.

Sep. 2021 - Jun. 2026 (Expected)

Visvesvaraya Technological University

Bachelor of Engineering in Computer Sciences - CGPA 8.16/10

Bangalore, KA, India

Aug. 2015 - Jun. 2019

Publications

- [1] Gupta, S., Biggs, S., Laber, M., **Shafi, Z.**, Walters, R., Paul, A. 2026. **DeepWeightFlow: Re-Basined Flow Matching for Generating Neural Network Weights** In The Fourteenth International Conference on Learning Representations (ICLR). [\[Link\]](#)
- [2] **Shafi, Z.** and Kadioglu, S., 2025. **FORGE: Foundational Optimization Representations from Graph Embeddings**. arXiv preprint arXiv:2508.20330 (In DiffCOAlg Workshop NeurIPS 2025). [\[Link\]](#)
- [3] **Shafi, Z.**, Savcicens, G., and Eliassi-Rad, T., 2025. **REGE: A Method for Incorporating Uncertainty in Graph Embeddings**. In Proceedings of the 2025 SIAM International Conference on Data Mining (SDM). [\[Link\]](#)
- [4] **Shafi, Z.**, Miller, B.A., Eliassi-Rad, T. and Caceres, R. S., 2025. **Accelerated Discovery of Set Cover Solutions via Graph Neural Networks**. In International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR). [\[Link\]](#)
- [5] **Shafi, Z.**, Chatterjee, A., and Eliassi-Rad, T., 2025. **Explaining Node Embeddings**. In Transactions on Machine Learning Research (TMLR). [\[Link\]](#)
- [6] Miller, B.A., **Shafi, Z.**, Ruml, W., Vorobeychik, Y., Eliassi-Rad, T. and Alfeld, S., 2025. **Defense Against Shortest Path Attacks**. In Proceedings of the 2025 SIAM International Conference on Data Mining (SDM). [\[Link\]](#)
- [7] Miller, B.A., **Shafi, Z.**, Ruml, W., Vorobeychik, Y., Eliassi-Rad, T. and Alfeld, S., 2023. **Attacking Shortest Paths by Cutting Edges**. ACM Trans. Knowl. Discov. Data (TKDD) 18, 2, Article 35 (February 2024), 42 pages. [\[Link\]](#)
- [8] **Shafi, Z.**, Miller, B.A., Chatterjee, A., Eliassi-Rad, T. and Caceres, R. S., 2023. **GRASP: Accelerating Shortest Path Attacks via Graph Attention**. In Deep Learning on Graphs Workshop, Knowledge Discovery and Data Mining (KDD) [\[Link\]](#)
- [9] Chatterjee, A., Walters, R., **Shafi, Z.**, Ahmed, O.S., Sebek, M., Gysi, D., Yu, R., Eliassi-Rad, T., Barabási, A.L. and Menichetti, G., 2023. **Improving the generalizability of protein-ligand binding predictions with AI-Bind**. Nature Communications, 14(1), p.1989. [\[Link\]](#)
- [10] Miller, B.A., **Shafi, Z.**, Ruml, W., Vorobeychik, Y., Eliassi-Rad, T. and Alfeld, S., 2021, September. **PATHATTACK: Attacking Shortest Paths in Complex Networks**. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD) (pp. 532-547). Springer. [\[Link\]](#)
- [11] Liu, D *, **Shafi, Z.***, Fleisher, W., Eliassi-Rad, T. and Alfeld, S., 2021, July. **RAWLSNET: Altering Bayesian Networks to Encode Rawlsian Fair Equality of Opportunity**. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES) (pp. 745-755). [\[Link\]](#)
- [12] Mishra, S., **Shafi, Z.** and Pathak, S. 2019. **Time Series Event Correlation with DTW and Hierarchical Clustering Methods**. PeerJ Preprints No. e27959v1. [\[Link\]](#)

Work Experience

Fidelity Investments

Boston, MA, U.S.

Co-op - Data Scientist - AI Center of Excellence

Jan. 2025 - Jul. 2025

- **Foundational Model for Mixed Integer Programs (MIPs)** [2].

Goal: Design, train, test and deploy a foundational model to represent general mixed integer programming instances

Impact: Created novel vector quantized graph-based representation approach that led to publication at NeurIPS 2025 DiffCOAlg Workshop (FORGE paper), enabling improved optimization solution strategies

Scale: General-purpose model applicable across diverse MIP problem classes

- **Customer Representation Architecture**

Goal: Design, train and test a foundational architecture for representing customer data to support ML applications like product and article recommendation. Compare against existing baseline models.

Impact: Developed reusable representation system that enhanced model development efficiency and consistency. Proposed model performed on par with existing models with ≈80% fewer input features.

Akamai Technologies

Performance Engineer II - Global Performance And Operations

Jul. 2019 - Jul. 2021

- **Platform Performance Optimization**

Goal: Optimize critical platform performance metrics to improve user experience and system efficiency

Impact: Enhanced system efficiency through measurable improvements in throughput, content offload ratios, and latency reduction on a per customer basis.

Scale: Global Akamai CDN platform. Offload scales were split across mapped memory hierarchies achieving up to 90% for most customers. Throughput scales ranged between 3Mbps - 11Mbps depending on customers and countries.

- **Log Data Mining System**

Goal: Design and implement scalable systems for extracting insights from raw server logs

Impact: Enabled data-driven decision-making through actionable insights, improving operational efficiency and issue resolution

Scale: Processing logs from distributed global server infrastructure - a few million logs per customer per hour

- **Critical Event Scaling**

Goal: Scale and prepare platform infrastructure for high-traffic events requiring exceptional performance

Impact: Successfully handled traffic for major global events including IPL cricket league and Apple WWDC keynote without degradation

Scale: Up to 10 Tbps of traffic in a single country

- **Intern Mentorship**

Goal: Mentor and guide junior engineers through complete project lifecycle

Impact: Ensured successful project completion and presentation delivery, developing next-generation engineering talent

Akamai Technologies

Intern - Platform & Delivery

Jan. 2019 - May 2019

- **Root Cause Analysis System [12]**

Goal: Develop automated system for identifying root causes of platform issues

Impact: Improved accuracy and speed of issue identification through time series correlation analysis, reducing mean time to resolution

Scale: Multiple parallel streams of time series data across distributed infrastructure

- **Network Traffic Visualization Tools**

Goal: Build visualization tools to understand traffic patterns and demand across the network

Impact: Improved capacity planning and resource allocation through granular traffic insights

Scale: Global Akamai network with drill-down capabilities from country level to individual server sets

Skills

Languages Python, Bash, JavaScript, C, SQL

Statistics Network Analysis, Anomaly Detection, Time Series Analysis, A/B Testing

Machine Learning NumPy, Pandas, SciKit-Learn, SciPy, NetworkX, iGraph

Deep Learning PyTorch, PyG, TensorFlow, Transformers

Visualization Matplotlib, Tableau, Seaborn, HighCharts, Grafana, Gephi, Cytoscape

Cloud Modal.ai, AWS, Azure, GCP

Academic Services

Reviewer KDD '24, ECAI '25, NeurIPS '25, AI Magazine '25, DiffCoALG@NeurIPS '25, NetSci '26

Program Committee IAAI '26

Instructor of Record DS 4400 - Machine Learning and Data Mining, Spring '26, Northeastern University [\[Course Homepage\]](#)