

Majid Daliri

370 Jay St, Brooklyn, NY
Floor 11th, Desk A1119

Cell: +1-646-750-4667
daliri.majid@nyu.edu
[majid-daliri.github.io](https://github.com/majid-daliri)

Education	New York University, New York, USA M.S. and Ph.D. in Computer Science Advised by Prof. Christopher Musco University of Tehran, Tehran, Iran B.Sc. in Computer Engineering (Cumulative GPA: 3.97/4.0)	2022 - 2026 2017 - 2022
Publications	<ul style="list-style-type: none">• Fast Vector Database Indexing for Efficient Passage Retrieval (Under Review) 2024 Majid Daliri, Amir Zandieh• Sampling Methods for Inner Product Sketching (VLDB) 2024 Majid Daliri, Juliana Freire, Christopher Musco, Aécio Santos, Haoxiang Zhang• Simple Analysis of Priority Sampling (SOSA) 2024 Majid Daliri, Juliana Freire, Christopher Musco, Aécio Santos, Haoxiang Zhang• KDEformer: Accelerating Transformers via Kernel Density Estimation (ICML) 2023 Amir Zandieh, Insu Han*, Majid Daliri*, Amin Karbasi (* equal contribution)• Weighted Minwise Hashing Beats Linear Sketching for Inner Product Estimation (PODS) 2023 Aline Bessa, Majid Daliri, Juliana Freire, Cameron Musco, Christopher Musco, Aécio Santos, Haoxiang Zhang• Efficient Approximations for Cache-conscious Data Placement (PLDI) 2022 Ali Ahmadi, Majid Daliri, Amir Kafshdar Goharshady, Andreas Pavlogiannis• A 10-Approximation of the $\frac{\pi}{2}$-MST (STACS) 2022 Ahmad Biniaz, Majid Daliri, AmirHossein Moradpour	
Selected Internship Program	Machine Learning Intern, Max-Planck-Institut für Informatik <ul style="list-style-type: none">• Implemented Fast Attention in Transformers, optimizing accuracy and efficiency in sequence modeling tasks.• Designed an efficient GPU-compatible LSH method, boosting performance in attention approximation.• Technical Stack: BigGAN, PyTorch, Transformer architectures, and advanced sequence modeling tools. Research Intern, HKUST <ul style="list-style-type: none">• Made a pioneering theoretical contribution to Cache-conscious Data Placement (CDP), addressing a long-standing challenge in optimizing cache hits.• Implemented and tested various cache management policies and algorithms, outperforming previous heuristics.• Our method emerged as the most effective solution, setting a new standard for cache optimization.	Apr 2021 - Jan 2022 Jun 2021 - Jun 2022
Selected Awards and Honors	Research Grant, University of Salzburg Awarded a €5,000 grant for a research internship focusing on algorithms for distribution bisimilarity, probabilistic systems verification, and quantum annealing projects. Hong Kong PhD Fellowship Scheme (HKPFS) scholarship totaling HK\$1,445,200 (approximately \$184,100). I was among the top 300 students selected worldwide across all majors, showcasing academic excellence and research potential.	Summer 2022 2022

	ACM ICPC - Regional (University of Tehran) ranked 6 th among more than 100 team all around the Iran.	2019
	Iranian National Olympiad in Informatics Finalist (IOI, Iran) are awarded to around 50 selected after a year of competition among over 10000 Students.	2016
Service	<ul style="list-style-type: none"> • Reviewer for International Conference on Machine Learning (ICML 2024) • Reviewer for Royal Society Open • External Reviewer for Canadian Conference on Computational Geometry (CCCG 2023) 	
Conference Presentations	<ul style="list-style-type: none"> • Simple Analysis of Priority Sampling Presentation, (SOSA) 2024 • Accelerating Transformers via Kernel Density Estimation Poster, (ICML) 2023 • Weighted MinHash for Inner Product Estimation Poster, (PODS) 2023 • Efficient Approximations for Cache-conscious Data Placement Presentation, (PLDI) 2022 	
Teaching	<ul style="list-style-type: none"> • Section Leader for CSCI-UA 310 Basic Algorithms Spring 2023 • Teaching Assistant NYU CS-GY 6763 Algorithmic Machine Learning Fall 2022 • Teaching Assistant UT Design and Analysis of Algorithms, H. Mahini Fall 2020-2021 	
Work Experience	Site Reliability Engineer at Cafebazaar 2021 - 2022 <ul style="list-style-type: none"> • Designed, implemented, and maintained both Redis-as-a-Service/PostgreSQL-as-a-Service on a Kubernetes-based cloud. • Achieved consistent performance benchmarks for both services with 100% uptime and a 99.9% response rate. • Technical Stack: Kubernetes, Docker, Sentry, S3, Prometheus 	
Skills and Qualities	Theoretical Background: Proficient in Machine Learning Theory, Neural Networks, Linear Algebra, and Probability. Technical Skills: Highly skilled in C/C++, Go, Python, Bash-Scripting, PHP, JavaScript. Experience with PyTorch, TensorFlow, Django, CSS3, HTML5, and git. Other Attributes: Innovative, self-driven, and communicative, with ability to work efficiently both independently and in a team.	