Dario Fumarola

New York, NY • dariofumarola90@gmail.com • +1 (540) 466-2013

Education

Washington and Lee University — Lexington, VA
Computer Science and Mathematics – Davis Scholar — *2019 – 2023*

- Relevant Courses: Deep Learning, ML and Big Data, Real Analysis, Network Security, Differential Geometry

Research Interests

My research develops mathematical frameworks to enhance the design, efficiency, and interpretability of deep learning systems. This includes advancing geometric techniques for high-dimensional vector spaces with applications in distributed retrieval, bioinformatics, and drug design. I also focus on using graph-theoretical models to improve structural integrity, explain complex behaviors, and establish principles that ensure geometric consistency across data representations.

Research Experience

Amazon Science — New York, NY
Research Assistant – Professor Hakan Ferhatosmanoglu — *2023 – Present*

- Led research on locality-sensitive hashing and graph clustering algorithms, improving distributed database queries
- Developed dimensionality reduction techniques combining DCT with learned quantization for vector embeddings
- Directed the implementation of a distributed vector search system, successfully replacing AWS legacy services
- Published two technical reports on ANNs search optimization, driving adoption across storage research divisions

Washington and Lee University — Lexington, VA
MATH 332 Differential Equations – Course Assistant — *2021*

- Conducted weekly review sessions for 40+ students, focusing on numerical methods and stability analysis in ODEs
- Authored comprehensive LaTeX solution guides covering existence theorems, series methods, and boundary values
- Designed and led advanced MATLAB workshops on numerical integration schemes and phase plane analysis
- Created interactive diagrams on applications of eigenvalue analysis in dynamical systems and control theory

Industry Experience

Amazon Web Services — New York, NY
Solutions Architect – Prototyping Team — *2023 – Present*

- Engineered cloud-based prototypes leveraging transformer architectures to advance bioinformatics workflows
- Developed and maintained a novel drug discovery platform integrating AlphaFold3 and ESM models
- Architected end-to-end ML pipelines for high-throughput genomic sequencing, reducing inference time by 60%
- Implemented attention-based RAG system for biomedical literature, achieving 90% recall across 10M+ papers

Certifications: AWS Cloud Practitioner; AWS Solutions Architect – Associate; AWS Machine Learning – Specialty
Professional Memberships: Institute of Electrical and Electronics Engineers; Association for Computing Machinery

Selected Projects

Hierarchically Partitioned Cloud-Native Vector Search
High-performance cloud indexing system, scheduled for future AWS product launch

- Built distributed vector search system achieving better throughput than DiskANN via novel S3 parallelization
- Developed hierarchical clustering algorithm based on HNSW graphs, with 90% recall at sub-second latency
- Implemented product quantization with adaptive codebook learning, reducing storage requirements by 75%

Selective Replication for Efficient k-NN Retrieval
Exploration of advanced vector duplication strategies in machine learning clustering

- Formulated theoretical proofs demonstrating the efficiency of topologically-aware replication over centroid methods
- Developed an adaptive replication strategy based on diversity estimation, reducing duplicates and improving recall
- Optimized vector distribution through R-NN modeling, decreasing need for vectors accessed by 15% at equal recall

Geometry-Enriched Graph Attention for Molecular Insights
Cloud-based deep learning framework for molecule analysis and interaction prediction

- Developed a Graph Attention Networks (GATs) for drug candidates screening, averaging 91% prediction accuracy
- Created custom geometric attention mechanisms for drugs and trained the model on a dataset of 100,000 molecules
- Presented findings at the Amazon HQ2 Conference, highlighting improvements over traditional screening methods