Dario Fumarola

New York, NY – dariofumarola90@gmail.com – (540) 466-2013

EDUCATION

Washington and Lee University Lexington, VA
Davis Scholar – Bachelor's Degree in Computer Science and Mathematics 2019 – 2023
Coursework: Deep Learning, Differential Equations, Real Analysis, Network Security, Topology

RESEARCH INTERESTS

- Using concepts from geometry, information theory, and topology to guide model design, ensuring more stable and interpretable neural representations.
- Employing graph neural networks and manifold learning to capture rich relational information in data, improving accuracy and transparency.
- Integrating distributed training, efficient architectures, and resource-aware approaches to tackle billion-scale datasets, accelerating innovation while maintaining model integrity.

RESEARCH EXPERIENCE

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

- Investigated vector embedding storage optimization techniques emphasizing geometric and topological integrity in billion-scale, high-dimensional datasets for large-scale retrieval
- Engineered an S3-based storage architecture leveraging partitioned graph structures to preserve underlying manifold characteristics and improve data organization
- Implemented density-aware replication strategies that maintained intrinsic data geometry, achieving a 3x improvement in recall at fixed access costs
- Authored technical reports and mentored a Ph.D. Applied Scientist from Cornell, collaboratively integrating geometric insights into system design and documentation

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

- Led weekly review sessions for 40+ mathematics major students, focusing on nonlinear ODEs, numerical methods for PDEs, and stability analysis
- Created LaTeX guides and interactive visualizations illustrating existence theorems, eigenvalue analysis, and control theory, linking fundamental math concepts to deep learning stability
- Developed MATLAB workshops on numerical integration and phase-plane analysis, using classical systems to highlight the geometric intuition underlying complex model dynamics
- Implemented a continuous-depth Neural Network as a final project, leveraging neural ODE frameworks to enhance time-series modeling, stability, and interpretability in deep learning

RESEARCH PAPERS

FUMAROLA, D., Ferhatosmanoglu, H. "Selective Vector Replication for Improved Recall and Query Performance of Approximate Nearest Neighbor Search." In Preparation at *International Conference on Very Large Data Bases – (VLDB)*, 2024.

INDUSTRY EXPERIENCE

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

- Engineered custom attention layers in sequence-to-sequence bioinformatics pipelines, leveraging structural patterns in protein sequences to enhance interpretability and achieve 85 % accuracy
- Integrated Graph Neural Networks with advanced protein language models, scaling to 100M+ compounds, to discover latent structural patterns in drug-target interactions
- Designed distributed genomic model pipelines using gradient checkpointing and model parallelism, accelerating inference by 60 % while preserving meaning in embedded biological data
- Built a retrieval-augmented generation system with dense, topologically aware embeddings and hierarchical transformers, attaining 90 % recall across vast biomedical corpora

Certifications: AWS Cloud Practitioner, AWS Solutions Architect, AWS Machine Learning

Professional Memberships:

- Institute of Electrical and Electronics Engineers (IEEE)
- Association for Computing Machinery (ACM)
- Association for the Advancement of Artificial Intelligence (AAAI)
- Out in Science, Technology, Engineering, and Mathematics (oSTEM)

PRESENTED PROJECTS

Hierarchically Partitioned Cloud-Native Vector Search

- Led research on a high-performance cloud-based vector indexing framework for billion-scale embeddings, achieving 2x throughput over DiskANN and ensuring stable retrieval
- Designed hierarchical clustering techniques extending HNSW with learned distance metrics and adaptive routing, reaching 90 % recall at sub-second latency
- Implemented product quantization with learned codebooks and residual embeddings, reducing storage by 75 % while preserving robust high-dimensional data representations.

Geometry-Enriched Graph Attention for Molecular Insights

- Introduced a geometric deep learning framework for drug discovery on molecular graphs, presented at the Amazon HQ2 Bioinformatics Conference and adopted by pharmaceutical partners
- Developed a custom Graph Attention Network with molecular-aware message passing, delivering 91 % accuracy in predicting drug solubility and binding affinity
- Built multi-scale attention mechanisms to capture both local atomic interactions and global molecular structure, training on 100,000+ SMILES molecules

Selective Replication for Efficient k-NN Retrieval

- Led theoretical research on topology-preserving vector replication as first author, constructing a mathematical framework for boundary vertex identification in high-dimensional embeddings
- Formulated a replication strategy integrating reverse nearest neighbors and density-sensitive boundary detection, outperforming SPANN and soft clustering methods by 3x
- Established formal proofs for optimal boundary vertex selection, providing theoretical guarantees that minimize replication overhead while safeguarding neighborhood integrity