Bowen Li

1465 Nena Hills Ct. – Tallahassee, Florida – 32304

☐ (850)-570-6618   ✉ li.bw0712@gmail.com   🌐 bowenli1994.github.io
⚲ github.com/BowenLi1994   in/bowenli94   🎓 Google Scholar

EDUCATION

Florida State University   **Tallahassee, FL**
Ph.D. in Computer Science, GPA: 4.0/4.0   *Aug. 2019 – May 2026 (**Expected**)*

Florida State University   **Tallahassee, FL**
M.S. in Computer Science, GPA: 3.8/4.0   *Aug. 2017 – May 2019*

East China University of Science and Technology   **Shanghai, China**
B.Eng. in Computer Science and Technology   *Sep. 2012 – Jun. 2016*

WORK EXPERIENCE

Data and Information Systems Laboratory (DAIS)   **Tallahassee, FL**
Research Assistant, Florida State University   *Aug 2019- Present*

Pursuing a Ph.D. under the supervision of Prof. Peixiang Zhao, focusing on scalable data mining and machine learning techniques for large-scale and high-dimensional data. Research areas include outlier detection, spectral clustering, and nearest neighbor search.

- **Outlier Detection**: Designed the oblique-subspace formulation and hashing architecture. The resulting **OS-Hash** algorithm breaks the limitations of prior methods that only operate on multidimensional data by extending outlier detection to arbitrary data types, including time-series and graph-structured datasets. It delivers strong empirical performance, achieving up to **10× faster runtime** and consistently higher detection recall on high-dimensional and schema-less datasets compared to state-of-the-art baselines.

- **Spectral Clustering**: Developed the stochastic factorization framework underlying **BSpec**, replacing the traditional $O(n^2)$ similarity-matrix construction with an efficient stochastic gradient optimization pipeline. This design removes the memory bottleneck of classical spectral clustering and enables scalable clustering with out-of-sample embedding support. BSpec demonstrates strong empirical improvements, achieving significantly faster computation and comparable or higher clustering quality on large-scale, high-dimensional datasets.

- **Approximate k-NN Search**: Designed a learning-based optimization framework for Hierarchical Navigable Small World (HNSW) graphs by tracking per-node and per-edge visitation frequencies across queries. Leveraging these learned statistics, reconstructed the upper-layer graph structure to guide future searches more efficiently, yielding faster convergence toward nearest neighbors and improved query performance.

Amazon Web Services (AWS)   **Seattle, WA**
Software Engineer Intern   *May 2023 – Aug 2023*

- Developed a centralized dashboard (**Opan Coral**) providing AWS RDS owners real-time visibility into API definitions and resource access for large multi-tenant RDS instances, significantly simplifying administration.

- Migrated the backend from ClojureScript to TypeScript, improving type safety and aligning the internal tool with modern AWS development standards for future scalability.

- Collaborated with UI/UX to design and implement a modernized interface that improved data readability and usability for both technical and non-technical users.

Amazon Web Services (AWS)   **Seattle, WA**
Software Engineer Intern   *May 2022 – Aug 2022*

- Engineered an asynchronous delete command for RDS Multi-AZ Outposts workflows, using DynamoDB as the command channel to cut up to 30 minutes of developer wait time per operation through on-demand termination.

- Improved workflow efficiency by adding a cancel feature for create workflows, preventing unnecessary compute cycles and resource use on long-running tasks.

- Designed a state-checking mechanism via a polling system integrated with DynamoDB, enabling on-demand control over active workflows.

Florida State University   **Tallahassee, FL**
Teaching Assistant   *Aug 2018 – Present*

- Promoted to **Lead Teaching Assistant** for Web Programming (CGS3066), leading curriculum design, creating all course materials, and delivering lectures to 40+ students.
- Mentored hundreds of students across 7 CS courses, improving code quality and strengthening understanding in C++, Algorithms, and Database Systems.
- Provided specialized instruction in Software Reverse Engineering (CAP5137), guiding students through hands-on labs and practical techniques.

PUBLICATIONS

Fast Outlier Detection in Oblique Subspaces   **CIKM' 25**
34th Conference on Information and Knowledge Management   *Nov. 2025, Seoul, Korea*

Proposed **OS-Hash**, a high-performance oblique-subspace hashing framework for detecting outliers in multidimensional, time-series, and graph data, delivering superior accuracy and significant speedups over prior methods.

Fast and Accurate Randomized Spectral Clustering   **ICDE '26**
42nd IEEE International Conference on Data Engineering   *under review*

Developed **BSpec**, a scalable stochastic spectral clustering algorithm that avoids $O(n^2)$ storage through on-the-fly similarity updates and enables efficient incremental and out-of-sample embeddings for large-scale systems.

SKILLS

- **Languages**: C/C++, Python, Shell Script, SQL, TypeScript, HTML/CSS, PHP
- **Expertise**: Data Mining, Machine Learning, Outlier Detection, Spectral Clustering, Nearest Neighbor Search
- **Technologies**: AWS (Lambda, Step Functions, DynamoDB), Git, Linux, Database Systems, Query Optimization