

Research Interests

Vector Databases, Information Retrieval, LLMs, NLP, Machine Learning Systems, AI.

Education

2022–Present **B.S. in Mathematics (With minor in Applied Mathematics)**, *University of Washington*, Seattle, WA, Dean's List at 2023, 2024, 2025
Coursework: Two-years Honors Math Sequence (Vector Calculus, Differential Equations, Real & Complex Analysis), Abstract Algebra, Topology & Geometry, Probability, DBMS, AI, and Machine Learning.

Publications

- ICDM 2025 **Jiuzhou Fu**, Luanzheng Guo, Nathan Tallent, Dongfang Zhao. *ProHD: Projection-Based Hausdorff Distance Approximation*. Oral Presentation at the **IEEE International Conference on Data Mining (ICDM) 2025** (top-tier data mining conference, 13% acceptance rate for full research paper). <https://arxiv.org/abs/2511.18207>
- Under review at ICDE **Jiuzhou Fu**, Dongfang Zhao. *QPAD: Quantile-Preserving Approximate Dimension Reduction for Nearest Neighbors Preservation in High-Dimensional Vector Search*. Under review at **IEEE International Conference on Data Engineering (ICDE) 2026**. <https://arxiv.org/abs/2504.16335>
- Under review at i2MTC Yuhan Huang, Xiaoxi Hu, **Jiuzhou Fu**, Fei Chen, Jingming Cao, Naibang Wang. *SSIDN: A Small-Sample Incremental Diagnosis Network for Railway Turnout Actuators*. Under review at **IEEE Instrumentation and Measurement Technology Conference (i2MTC) 2026** (The flagship conference of the IEEE Instrumentation and Measurement Society). [paper](#)
- Under review at i2MTC Xu Zhang, Xiaoxi Hu, Yuhan Huang, **Jiuzhou Fu**, Jingming Cao, Tao Tang. *Channel-Incremental Convolutional Neural Network for Railway Point Machine Fault Diagnosis with Expanding Sensor Configurations*. Under review at **i2MTC 2026**. [paper](#)

Award

ICDM 2025 Student Travel Award (Funded by NSF)

Research

- Nov 2024 – Present **Undergraduate Researcher**
High-Performance Data-Intelligence Computing (HPDIC) Lab, *University of Washington*, Seattle, Principal: Prof. Dongfang Zhao
- **ProHD: Projection-Based Hausdorff Distance Approximation (ICDM 2025)** — proposed a projection-guided subset-selection algorithm combining centroid intersections and principal components to approximate the Hausdorff distance up to 10–100× faster with bounded error; work supported by the **U.S. Department of Energy's Pacific Northwest National Laboratory (PNNL)**, Texas A&M ACES cluster (NSF ACCESS), and Chameleon testbed; presented in Washington, D.C.
 - **QPAD: Quantile-Preserving Approximate Dimension Reduction for Nearest Neighbors Preservation in High-Dimensional Vector Search (under ICDE review)** — developed an unsupervised dimensionality-reduction method maximizing neighborhood margin to preserve approximate nearest-neighbor (ANN) structure, with formal proofs of boundedness, compactness, and indexability in ANN systems.
 - **Directed Hypergraph Compression and Reordering (under preparation)** — extended a state-of-the-art undirected hypergraph reordering and compression algorithm to directed hypergraphs, enabling dynamic maintenance, reduced L1 cache misses, task speedups, and a smaller storage footprint for large-scale graph systems.
 - **Techniques and Stack:** **Linux/VM** environment; **C/C++** (*STL*, *OpenMP*, *Makefile*); **Python** (*multiprocessing*, *psutil*, *pandas*, *NumPy*, *SciPy*); *FAISS* (ANN search); *matplotlib* (visualization); *mpi4py* and *PyOpenCL* (parallel computation); *Conda* (environment management).

Mar 2025 – **Paid Research Intern**

Present **School of Vehicle and Mobility, Tsinghua University, Beijing**

- **SSIDN: A Small-Sample Incremental Diagnosis Network for Railway Turnout Actuators (under i2MTC review)** — Simulated time-series sensor-anomaly datasets (open-circuit, periodic interference, impulse noise, etc.) for railway switch machine and industrial robots to evaluate fault-diagnosis robustness, enabling end-to-end classification under realistic sensor failures.
- **CI-CNN: Channel-Incremental Convolutional Neural Network for Railway Point Machine Fault Diagnosis (under i2MTC review)** — We developed a channel-incremental framework that integrates new sensors without retraining via orthogonal initialization and selective freezing; validated on ZDJ9-RPM deployments, achieving 0.945 accuracy with minimal overhead.
- Implemented data preprocessing and **error-injection pipelines** in **Python** (*NumPy, pandas, PyTorch*) for vibration and control-signal time series; performed ablation studies and feature visualizations.

Dec 2023 – **Undergraduate Researcher**

Dec 2024 **Washington Experimental Mathematics Lab (WXML), University of Washington, Seattle**

- Conducted theoretical and computational research on the **Concentration of Measure** phenomenon via the Curie–Weiss model and high-dimensional spheres, deriving analytic asymptotics and probabilistic limits, and validating them through **Monte Carlo simulations** and symbolic computation in *Wolfram Mathematica*.
- Presented the project poster at the **UW Undergraduate Research Symposium** as part of the **Undergraduate Research Program**, highlighting rigorous mathematical reasoning combined with computer-aided experimentation. [Poster link](#)

Internship

Jun – Aug 2025 **Prompt Engineer Intern, MaaS Department, Z.AI, Beijing, China**

- Worked on the **REDNOTE** (Xiaohongshu overseas version) AI Translation System Team, focusing on multilingual prompt design, retrieval-augmented generation (**RAG**) pipelines, and translation strategy development across cross-lingual domains.
- Designed and deployed **Dify-based prompt-engineering workflows**, supporting **instruction fine-tuning** of multimodal LLMs (GLM-4v/4.5v) for contract understanding from PDF/image inputs and structured JSON output generation.
- Designed the **PRD** and **MVP** for an LLM prompt-sharing website; implemented workflows for prompt-to-HTML rendering, multi-speaker podcast generation, and TTS stress testing for production readiness.

May – Aug 2024 **Data Science Intern, Real AI, Beijing, China**

- Processed large-scale unstructured web data using **Python** (*pandas, NumPy*) and **SQL**, including automated data-cleaning pipelines for schema standardization, null-value imputation, and cross-table validation.
- Helped design and deploy a multi-turn prompt-engineering and retrieval pipeline for an enterprise policy-advising LLM, integrating natural-language understanding, semantic retrieval, and consistency checks to enhance factual reliability and reduce hallucination.

Aug – Oct 2024 **Cybersecurity Intern, Security BU, QAX (Qi An Xin Group), Beijing, China**

- Participated as a **Blue-Team Cybersecurity Junior Specialist** in red-blue adversarial exercises, performing real-time threat detection, vulnerability assessment, and network forensics using QAX's defense platform.
- Utilized *Burp Suite, sqlmap, AWVS*, and *Kali Linux* to identify injection, authentication, and configuration vulnerabilities; wrote technical reports and mitigation documentation for incident response.

Recommenders

Assistant **Prof. Dongfang Zhao, dzhao@uw.edu**, University of Washington

Professor Principal Investigator of the HPDIC Lab and my primary research mentor. Worked together on three projects involving high-performance data-intelligence systems.

Computer **Dr. Luanzheng Guo, lenny.guo@pnnl.gov**, Pacific Northwest National Laboratory

Scientist Collaborator on power-consumption analysis for data centers and co-author of ProHD: Projection-Based Hausdorff Distance Approximation.

Postdoctoral **Dr. Hadrian Quan, hquan1@ucsc.edu**, University of California, Santa Cruz (formerly University of Washington)

Instructor for the year-long Accelerated Honors Advanced Calculus sequence and WXML faculty mentor for a nine-month mathematics project on concentration of measures.