

XUEYING DING

+1(901)-355-8092 | xding2@andrew.cmu.edu | Google Scholar

EDUCATION

Ph.D. in Public Policy and Machine Learning

Aug 2021 – Dec 2026 (Expected)

Carnegie Mellon University, Pittsburgh, PA

Co-advisors: Professor Leman Akoglu and Professor Pradeep Ravikumar

M.S. in Machine Learning

Aug 2019 – Dec 2020

Carnegie Mellon University, Pittsburgh, PA

B.S. in Computer Science and Mathematics

Aug 2015 – May 2019

Vanderbilt University, Nashville, TN

PROFESSIONAL EXPERIENCE

Research Intern at UMAP, Snap Research

May 2025 – Current

Supervisor: Tong Zhao, Neil Shah

Bellevue, WA

- Studied **decoder-only LLMs** as general-purpose embedding models and identified key limitations from causal attention, including restricted backward information flow and long-context information oversquashing.
- Proposed **Hierarchical Token Prepending (HTP)**, a training-free and architecture-agnostic method that introduces block-level summary tokens to improve information flow in LLM embeddings without modifying model weights. HTP improves both zero-shot and fine-tuned LLM embeddings across 11 retrieval datasets and 30 general embedding benchmarks, with at least 5% overall gains in information retrieval.
- Investigated **Reinforcement Learning with Verifiable Reward (RLVR)-based search agents** and **LLM-driven query rewriting strategies** for open-domain information retrieval and user-query based recommendation tasks.
- Collaborated on **Threshold Differential Attention (TDA)**, a sink-free, ultra-sparse attention mechanism for long-context LLMs that achieves > 99% exact zeros while matching Softmax attention on QA benchmarks. Pre-trained TDA LLMs improve SCROLLS passkey retrieval by 2.5x at 4k tokens.

Quant Research Intern at Bloomberg

May 2024 – Aug 2024

Supervisor: Achintya Gopal, Arun Verma

New York City, NY

- Developed **Delphyne**, a novel finance-oriented time-series foundation model inspired by LLM-style pre-training and adaptation, extending foundation-model design to continuous, noisy, multi-frequency financial signals.
- Investigated negative transfer in time-series pre-training and designed finance-aware pre-training/adaptation strategies using large-scale public time-series data together with revenue, bar, ticker, return, and volatility data.
- Demonstrated strong downstream performance across stock-return prediction, log-volume forecasting, volatility modeling, and broader forecasting benchmarks, including finance-specific tasks as well as in-distribution Monash and out-of-distribution ETT settings.

Research Intern at Capital One

June 2022 – Aug 2022

Supervisor: Nikita Seleznev, Bayan Bruss, Philip Spiegel

McLean, Virginia

- Developed **explainable XStream**, an unsupervised fraud detection algorithm for high-dimensional streaming bank data, connecting anomaly detection with emerging LLM-style interpretability and reasoning workflows.
- Integrated feature-importance inference, counterfactual reasoning, XPACS group anomaly explanations, and Look-out pairwise plots to produce structured, human-readable explanations for complex fraud patterns.
- Built a prototype human-in-the-loop GUI that supports analyst investigation, feedback collection, and interactive reasoning over detected anomalies.

RESEARCH PUBLICATION

LLM Evaluation, Pre-/Post- Training, Foundation Models

X. Ding, H. Wen, S. Klüttermann, L. Akoglu, [From Zero to Hero: Advancing Zero-Shot Foundation Models for Tabular Outlier Detection](#), ICML 2026

X. Ding*, X.Huang*, M. Ju, L. Collins, Y. Liu, L. Akoglu, N. Shah, T.Zhao, [Hierarchical Token Prepending: Enhancing Information Flow in Decoder-based LLM Embeddings](#), ACL Main Track (Oral) 2026

X.Huang*, **X. Ding***, M. Ju, Y. Liu, N. Shah, T.Zhao, [Threshold Differential Attention for Sink-Free Ultra-Sparse, and Non-Dispersive Language Modeling](#), ACL Main Track (Oral) 2026

Y. Li, Y. Miao, **X. Ding**, R. Krishnan, R. Padman, [Firm or fickle? Evaluating Large Language Models Consistency in Sequential Interactions](#), ACL Findings 2025

X. Ding, L. Akoglu, [Toward Privileged Foundation Models: LUPI for Accelerated and Improved Learning](#), Under Review, 2026

X. Ding, A. Mittal, A. Gopal, [DELPHYNE: A Pre-Trained Model for Financial Time Series](#), Runner-up Best Paper Award at NeurIPS GenAI in Finance Workshop 2025

L. Zhao, **X. Ding**, L. Akoglu, [Pard: Permutation-Invariant Autoregressive Diffusion for Graph Generation](#), NeurIPS 2024

L. Zhao*, **X. Ding***, L. Yu, L. Akoglu, [Improving and Unifying Discrete & Continuous-time Discrete Denoising Diffusion](#), JMLR 2025

Outlier Detection, Hyperparameter Optimization, Dataset Curations and Others

X. Ding, S. Klüttermann, H. Wen, Y. Chen, L. Akoglu, [MacrOData: New Benchmarks of Thousands of Datasets for Tabular Outlier Detection](#), KDD 2026

X. Ding, R. Xi, L. Akoglu, [Outlier Detection Bias Busted: Understanding Sources of Algorithmic Bias through Data-centric Factors](#), AAAI (AI, Ethics, and Society), 2024

Y. Qin, Y. Zhang, Y. Nian, **X. Ding**, Y. Zhao, [MetaOOD: Automatic Selection of OOD Detection Models](#), ICLR 2025

X. Ding, Y. Zhao, L. Akoglu, [Fast Unsupervised Deep Outlier Model Selection with Hypernetworks](#), KDD 2024

Z. Zhao, **X. Ding**, Y. Zhao, BA Prakash. [PINNsFormer: A Transformer-Based Framework For Physics-Informed Neural Networks](#), ICLR 2024

X. Ding, N. Seleznev, S. Kumar, CB. Bruss, L. Akoglu, [From Explanation to Action: An End-to-End Human-in-the-loop Framework for Anomaly Reasoning and Management](#), ACM International Conference on AI in Finance (ICAIF) (Oral) 2023

X. Ding, L. Zhao, L. Akoglu, [Hyperparameter Sensitivity in Deep Outlier Detection: Analysis and a Scalable Hyper-Ensemble Solution](#), NeurIPS 2022

K. Liu, Y. Dou, **X. Ding**, X. Hu, R. Zhang, H. Peng, L. Sun, P. Yu, [Pygod: A python library for graph outlier detection](#), JMLR 2024

AWARDS

- Gold Reviewer Award, ICML 2026 (Top 25%)
- Oral Presentation, ACL 2026
- Best Paper Award Runner-up, NeurIPS GenAI in Finance Workshop 2025
- Oral Presentation, ACM International Conference on AI in Finance (ICAIF) 2023

SERVICE

Reviewers: TMLR, ICML, KDD, Neurips, ICLR, ECML, AAAI

Invited Talks:

- From Zero to Hero: Advancing Zero-Shot Foundation Models for Tabular Outlier Detection; PriorLab; February 2026
- Outlier Detection Without the Pain: From Hyperparameter Sensitivity to Foundation Models; Snap Inc.; February 2026

Organizers: Workshop on Graph Foundation Models: A New Era of Graph Machine Learning, ICML 2026.