Andrew Koulogeorge

andrewkoulogeorge@gmail.com • 847-204-3005 • Personal Website • Github • linkedin

Education

Carnegie Mellon University — Pittsburgh, PA | December 2025
- **Masters of Science**: Computer Science — GPA: 4.00/4.00
- **Coursework**: Machine Learning, Convex Optimization, Mathematical Statistics, Distributed Systems

Dartmouth College — Hanover, NH | June 2024
- **Bachelor of Arts**: Major in Mathematics, Minor in Computer Science — GPA: 3.98/4.00
- **Awards:** Summa Cum Laude, Phi Beta Kappa

Experience

AppLovin — Palo Alto, CA | May 2025 – August 2025
Research Scientist Intern
- Researched methods to improve AppLovin's bid prediction for impressions on the MAX auction house
- Adapted the *Implicit Quantile Network* (IQN) from the Reinforcement Learning community to estimate an auction's "Minimum Bid to Win" distribution
- Implemented, trained, and evaluated the IQN model on historical auction data and conducted A/B testing. Launched the IQN model on both the iOS and Android platforms, contributing **~$45 million/year to AppLovin's margin** with predictions **touching 1.2 billion daily users**

Forge Lab — Pittsburgh, PA | March 2025 – Present
Machine Learning Researcher: Advised by Virginia Smith
- Efficient Inference for Vision Language Models via leveraging attention sinks in Self-Attention Mechanism

Harpin AI — Bend, OR | June 2024 – August 2024
Applied Scientist Intern
- Applied text embedding models to enhance Harpin's core profile similarity model; trained an XGBoost classifier on over 100k record pairs, achieving a 1.5% improvement in model F1 score
- Pioneered the development of a Siamese Neural Network-based profile similarity model to enable Harpin to bypass expert feature creation and frictionlessly target customer use cases outside of identity data

Minds, Machines, and Society Lab — Hanover, NH | March 2023 – June 2024
Machine Learning Researcher: Advised by Soroush Vosoughi
- Identified fundamental flaws in interpretable Large Language Model architectures which resulted in unfaithful model explanations; proposed the *Faithful Alignment* framework to restore faithful model explanations
- Implemented the *Faithful Alignment* framework in PyTorch and demonstrated that it maintains strong model performance across various Natural Language Processing tasks

Publications

A. Koulogeorge, S. Xie, S. Hassanpour, S. Vosoughi
Bridging the Faithfulness Gap in Prototypical Models. *Insights Workshop; NAACL 2025 (Oral Presentation)*

W. Ma, H. Scheible, B. Wang, G. Veeramachaneni, P. Chowdhary, A. Sun, A. Koulogeorge, L. Wang, S. Vosoughi.
Deciphering Stereotypes in Pre-Trained Language Models. *2023 EMNLP*

Select Projects

Needle — Code | Python, Cuda | December 2024 – January 2025
- Built PyTorch inspired Deep Learning framework that supports AutoDiff, common Neural Network layers, Optimizers and Datasets/Data-loaders. Implemented backend operations in Cuda for GPU support

Distributed Systems — Code | C, Java | January 2025 – April 2025
- Implemented Remote Procedure Calls (RPCs) for Linux file operations, Distributed File-Caching Proxy with Session Semantics, Dynamically Scalable Web Service policy, and Two-Phase Commit (2PC)

Skills

Languages: Python, C, Java, Cuda
Libraries & Tools: PyTorch, Weights & Biases, Hugging Face, Numpy, XGBoost, Scikit-learn, Git, Pandas