ANDREW KOULOGEORGE

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

EDUCATION

Carnegie Mellon University Pittsburgh, PA | May 2026
- **Masters of Science**: Computer Science GPA: 4.00/4.00
- **Coursework**: Inference Algorithms for LLMs (audit), 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

Pinterest Remote | September 2025 – December 2025
Machine Learning Engineering Intern
- **Advanced Technology Group (ATG)** working on **Foundation Models for Recommendation Systems**

Forge Lab Pittsburgh, PA | March 2025 – Present
Machine Learning Researcher: Advised by Virginia Smith
- Exploring **Efficient Inference for Large Language Models** via **Attention Sinks** in Self-Attention Mechanism

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**

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

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/C++, Java, Cuda
Libraries & Tools: PyTorch, Weights & Biases, Hugging Face, Git, Numpy, Pandas, XGBoost, Scikit-learn
Cloud: GCP & AWS