# ANDREW KOULOGEORGE

akouloge@andrew.cmu.edu • 847-204-3005 • Personal Website • Github • linkedin

#### **EDUCATION**

#### **Carnegie Mellon University**

- Masters of Science: Computer Science
- Coursework in Progress: Machine Learning (PhD), Mathematical Statistics, Introduction to Computer Systems

# **Dartmouth College**

- Bachelor of Arts: Major in Mathematics, Minor in Computer Science
- Awards: Summa Cum Laude, Phi Beta Kappa
- Relevant Coursework: Randomized Algorithms (Grad), Deep Learning Robustness, Machine Learning, Artificial Intelligence, Object Oriented Programming, Real Analysis (Honors), Linear Algebra, Econometrics

### **PROFESSIONAL EXPERIENCE**

# Harpin AI

Applied Scientist Intern

• Researched the application of 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

Large Language Model 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
- Applied prompt engineering to Chat-GPT to construct The ImplicitStereo Dataset, a novel dataset used to investigate stereotype representation and detection in pre-trained Large Language Models

# **Dartmouth Computer Science Department**

Undergraduate Teaching Assistant

• Assisted in the instruction of COSC 74: Machine Learning and COSC 78: Deep Learning by providing one-onone tutoring and bi-weekly office hours to review theory, PyTorch, Scikit-learn and Numpy concepts

# **PUBLICATIONS**

A. Koulogeorge, S. Xie, S. Hassanpour, S. Vosoughi

Is Faithfulness in Prototypical Language Models a Mirage? Submitted to 2024 EMNLP

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. Accepted to 2023 EMNLP (paper)

### **SKILLS**

Languages: Python, C. SOL, Java Frameworks & Libraries: PyTorch, Hugging Face, Scikit-learn, XGBoost, Pandas, Numpy, Apache Spark, Git Amazon Web Services: SageMaker, S3, Bedrock

### **SELECT PROJECTS**

**Reinforcement Learning** 

Code | Stanford | January 2024 – March 2024 Self-studied Stanford CS234; implemented Value Iteration for finite state Markov Decision Process, Deep Mind's Deep O-Network to play Atari, and Proximal Policy Optimization for Cart-pole & Pendulum simulation

# **Transformer from Scratch**

Code | Personal Project | December 2023 • Implemented decoder only Transformer architecture inspired by seminal "Attention is All You Need" paper from scratch in PyTorch; trained model on GPU hardware to generate Harry Potter inspired text

### **Traditional AI**

Code | Dartmouth | September 2023 – November 2023 • Implemented A\* Search and the Viterbi Algorithm for Blind Robot Planning, Minimax + Alpha-Beta Pruning for Chess AI, WalkSAT for Sudoko solver, and Probabilistic Roadmap for Robotic Arm Planning

Hanover, NH | March 2023 – June 2024

Bend, OR | June 2024 – August 2024

Hanover, NH | January 2024 – June 2024

Pittsburgh, PA | December 2025

Hanover, NH | June 2024 GPA:3.98/4.00