# Rohan Jha

rjha@andrew.cmu.edu • robro612@gmail.com • 858-229-3216

San Diego, California • Pittsburgh, Pennsylvania

## EDUCATION

**Carnegie Mellon University**                                                                                    **May, 2023**

*B.S. Artificial Intelligence*                                                                                           *Pittsburgh, PA*

- GPA: 3.89, Dean's List High Honors
- Relevant Coursework: Parallel and Sequential DS and Algorithms (15-210), Introduction to Machine Learning (10-315), Search Engines (11-442), Introduction to Deep Learning (11-485), Convex Optimization (10-725)

## SKILLS & INTERESTS

- **Languages:** Python, C, Java, Standard ML, Bash
- **Technologies:** Git, Mercurial, NumPy, PyTorch, LaTeX, Junit & Mockito, AWS, Slurm, Conda, W&B
- **Concepts:** Optimization, Machine Learning, Data Structures & Algorithms, Functional Software Development
- **Interests:** Information Retrieval, Representation Learning, AI Robustness & Explainability

## EXPERIENCE

**Software Engineering Intern**                                                           **May, 2022 – August, 2022**

*Meta (Ads Core ML Modeling Intelligence)*                                                              *Menlo Park, CA*

- Using Caffe2/PyTorch, implemented sparse Mixture of Experts and novel DSelectK Gating techniques into multiple sub-architectures of production advertisement recommendation models
- Conducted validation and ablation experiments to determine the efficacy and infrastructure costs of newly introduced modules, achieving model performance improvements supporting multiple organizations across the company corresponding to increased revenue

**Research Assistant**                                                                     **December, 2021 – Present**

*Carnegie Mellon University (Language Technology Institute)*                                             *Pittsburgh, PA*

- Supported grad-level research in neural information retrieval focused on combining dense language models' context with the sparse efficiency of the inverted list architecture
- Designed, coded, and presented experiments and results to the principal investigator
- Produced conference-style work analyzing the performance/cost tradeoff and distribution of sparsely-factorized dense embeddings as a performance-preserving, cost-optimizing modification to an existing retriever

**Teaching Assistant**                                                                      **January, 2021 – May, 2022**

*Carnegie Mellon University (15-281: AI: Representation and Problem Solving, 07-180: Concepts of AI)*     *Pittsburgh, PA*

- Designed, tested, proctored, and graded written and programming-based homework and exams
- Led 120+ students with weekly office hours, recitations, and exam review sessions
- Was primary expert for student questions on class forum with high coverage and low response latency

**Software Development Engineer Intern**                                              **June, 2021 – August, 2021**

*Amazon (Prime Video Ads)*                                                                                     *Seattle, WA*

- Designed, implemented, and tested software processing billions of customer impression records each hour
- Led extensive design and review process after researching and documenting various solution alternatives
- Reduced latency, cost, and on-call pain points to accommodate imminent 3-7x increase in impression volume