Vaibhav Gupta

Mobile: 646-645-6284 | **Email:** vvg239@nyu.edu | **Website:** vaibhavgupta.io | **GitHub:** guptv93

EDUCATION

Courant, New York University — New York, USA
Master of Science in Computer Science; GPA: 3.97/4 — *Sep. 2018 – May 2020*

Dhirubhai Ambani Institute of Information and Communication Technology — Gujarat, INDIA
Bachelors of Technology in Computer Science; GPA: 8.0/10 — *Aug. 2010 – June 2014*

EXPERIENCE

LinkedIn Corp — Sunnyvale, USA
Machine Learning Engineer — *Jan. 2022 – Present*
- Working with the AI Foundations team, to develop state of the art optimization techniques for language models and recommendation systems.
- Improving performance metrics (training time and AUC) for Feeds and Ads deep-learning models at LinkedIn.
- Working on techniques like Explore Exploit (using Thompson Sampling) for Neural Networks.

Amazon — Seattle, USA
Applied Scientist II — *Feb. 2021 – Dec. 2021*
- Worked with the Seller Insights Team, helping Amazon sellers make better business decisions by providing them data driven insights.
- Built causal inference models to help sellers understand how much downstream impact each of their actions (like creating sponsored ads, or restocking inventory) would have in the long-term.

Amazon — Bangalore, INDIA
Software Development Engineer — *July 2014 – August 2017*
- Worked on a website product end-to-end, from Frontend UI to Backend Services and data-stores.
- Later worked with the Social Ads Team on Revenue Optimization and Big Data Analytics.

PAPERS

Self-Supervised Learning through the Eyes of a Child — **Published at NeurIPS 2020**
- Utilized modern self-supervised deep learning methods and a recent longitudinal, egocentric video dataset recorded from the perspective of several young children, to model cognitive development in human babies.
- ArXiv Link: *2007.16189* | Press Coverage - *NewScientist*, *DigitalTrends*

PROJECTS

Explore Exploit for Neural Networks — May 2023 – Present
- Developed a TF wrapper, to calculate covariance matrix (inverse of loss hessian) of selected model parameters. Using that we calculate variance of output logits and return a thompson sample for explore-exploit.
- ArXiv paper link for more details: *1912.00832*

Model Training Speedup for Ads AI and Job Recommender — Aug 2022 – Dec 2022
- Used techniques like Gradient Norm Clipping, Linear Warmup, LR Decay etc. to train with larger batch size without dropping AUC.
- In addition, used techniques like XLA and Mixed Precision Training to speed-up training.

Sparse Feature Modeling for Feed SPR — Mar 2022 – Aug 2022
- Used various techniques, like different embedding sizes, model architectures, optimization techniques, activation functions, etc. to use object ID feature for improving Feed model AUC.
- Tried one epoch training to reduce sparse feature overfitting. Paper link: *Towards Understanding Overfitting of Deep CTR*

PROGRAMMING SKILLS

Languages: Python, Java, JavaScript, CUDA   **ML Libraries:** Tensorflow, NumPy, PyTorch, Scikit-learn