

# Anshul Wadhawan

[Website](#) | [LinkedIn](#) | [Google Scholar](#) | anshulw@alumni.upenn.edu | New York City, NY | 267-938-0707

## EDUCATION

---

**University of Pennsylvania** Philadelphia, PA  
*Master of Science in Engineering in Computer and Information Science; GPA: 3.8/4.0 Aug 2021 - May 2023*

**Netaji Subhas University of Technology, Delhi University** Delhi, India  
*Bachelor of Engineering in Computer Engineering; GPA: 3.75/4.0 Aug 2016 - Aug 2020*

## EXPERIENCE

---

**Bloomberg LP** New York City, NY  
*Machine Learning Engineer* May 2023 - Present

- Contributed to the successful release of a scalable real-time Automatic Speech Recognition (ASR) system for processing streaming audio data, demonstrating proficiency in the development of high-performance data pipelines.
- Successfully integrated a Kubernetes based workflow to automate ASR model evaluations, achieving a ~3.3x speedup.
- Enhanced text punctuation and capitalization model, achieving a ~13% increase in F-score through iterative improvements, while also conducting comprehensive data analysis and refining the training set creation methodology.

**Quora** Mountain View, CA  
*Software Engineer - Machine Learning Intern* May 2022 - Aug 2022

- Independently developed and deployed pipeline in Python that generates and bans clusters of fake users who upvoted spam content and have matching jaccard similarity using Locality Sensitive Hashing.
- Automated daily execution of Airflow DAGs that auto banned 16.7k users on first run, stabilizing to 3k daily bans which constituted 33.6% of all bans - second highest of all banning pipelines.
- Identified and engineered discerning features to filter false positives in the pool of users shortlisted for banning.
- Achieved 0.6% annualized upper bound on FPR and abated spam report volumes by 6.1%.
- Independently implemented Data Efficient Image Transformer to devise the first ever image score in Quora which aids in ranking image content on the user feed.
- Designed and developed dirty chains to automatically populate the score for every answer/post being created/edited.
- Delivered a scalable feature that enables users to listen to Quora answers in a week long hackathon collaboratively.

**Flipkart (Walmart Group)** Bangalore, India  
*Software Engineer - Machine Learning* Aug 2020 - Jul 2021

- Collaborated in an Agile team to build a CTC based hierarchical ASR model for Flipkart voice-based search system.
- Investigated impact of subword representations, language modelling, beam rescoring, layer normalization, compound scaling leading to ~32% reduction in relative WERs.
- Incorporated 35M unlabeled audios to decrease requirement of labeled audios by ~70%, curtailing training time by ~23%
- Spearheaded the transition to streaming the ASR model to increase user engagement time by ~15%.
- Organised bi-weekly paper presentations for circulation of latest research developments in ASR.

## SELECTED PROJECTS

---

- **Ingredient Replacement System - Alexa Taskbot Challenge** [LINK]: Implemented an ingredient replacement module that recommends substitutes for ingredients that the user does not have. The system finds K nearest neighbors for a particular ingredient in the FoodBERT embedding space, and shortlists them on the basis of a thresholding criteria. Dec 2022
- **Informativity Detection System - Published in EMNLP WNUT 2020** [LINK]: The system identifies whether a COVID-19 tweet is informative or not. Employed BERT-large model pre-trained on 22.5M COVID-19 tweets. Formulated ensemble learning by fine tuning on 7 subsets of labelled data, resulting in an increase of **0.8%** in F1-score. Jul 2020
- **Dynamic HTTP Server with Concurrent Processing**: Built a high-performance HTTP server in Java that can efficiently serve static files from a directory; can handle multiple, persistent connections concurrently and can process a variety of headers in any order; has low CPU usage when idle and can handle up to 17k requests per second. Implemented features of the Spark Framework API like dynamic content support; route matching, path and query parameters from scratch. Feb 2023

## TECHNICAL SKILLS

---

**Languages:** Python, Java, JavaScript, SQL, C++, HTML, CSS

**Frameworks and other tools:** PyTorch, Pandas, HIVE, Scikit, Git, Numpy, Github, OpenCV, TensorFlow, Keras, NodeJS, Bootstrap, Linux(Ubuntu), Apache Spark, Airflow, Web development and technologies, Object Oriented Programming, MySQL, Familiarity with Distributed Systems, Recommender Systems, Data Analysis and Visualization, Applied ML.

**Relevant Coursework:** Machine Learning, Deep Learning, Computational Linguistics (Natural Language Processing), Interactive Fiction and Text Generation, Analysis of Algorithms, Artificial Intelligence, Software Foundations

## ACHIEVEMENTS

---

- **Global Rank 1, 6** at WNUT2020 (EMNLP Workshop) Shared Task 1, 2.
- **Global Rank 361** at Codechef Cookoff 2019.
- **Global Ranks 2, 2, 4, 4** at WANLP2021 (EACL Workshop) Shared Task 1. (4 Subtasks)
- **National Rank 5** at Flipkart GRID Challenge. (Out of 6737 teams registered)
- **National Rank 3** at Infineon Data Science Hackathon. (Out of 312 teams registered)
- **National Rank 1** at GeeksForGeeks Weekly. (For two consecutive weeks)