

# Ashutosh Tiwari

669-292-7534 | [findashutoshtiwari@gmail.com](mailto:findashutoshtiwari@gmail.com) | [Linkedin@ashutosh-tiwari/](https://www.linkedin.com/in/ashutosh-tiwari/) | [Homepage/Portfolio](#) | [Github@thunderock](#)

## WORK EXPERIENCE

### Machine Learning Engineer 4 (ML Platform)

ADOBE FIREFLY

Jul. 2024 – Present  
*GENERATIVE AI, COMPUTER VISION, LARGE LANGUAGE MODELS*

- Currently part of ML Platform, focused on offline inference for videos and images, and feature store.
- Part of a small team that wrote our first version of our inference framework using Pytorch Lightning.
- Worked on our first version of our feature store that supports training large machine learning models trained on billions of images and videos.
- Wrote the first data quality framework used to ensure the quality of feature generation using Cerberus. This is used across enrichment pipelines to ensure the correctness of output.
- Designed distributed inference frameworks: online (vLLM + Ray for LLMs like Falcon-40B, Llama 2 70B) and offline (Ray Data + PyTorch Lightning achieving 10-50x faster image loading, 3-8K images/sec on 32 GPUs).
- Leading enrichment service with fire and forget pipelines with dataset sizes of billions of images and videos.

### Senior Software Engineer (Data Platform)

EVOLUTIONIQ

Sep. 2023 – Jul. 2024  
*DATA PLATFORM, NATURAL LANGUAGE PROCESSING, REGRESSION*

- Part of Scoring and Quality sub-team, working at the intersection of generative AI, fin-tech, and health sector.
- Led the development of data pipelines that ingest, and process data to be ingested, and used for training by our machine learning models.
- Led the development of pipelines that train our models to predict a claimant's expected time to return to work, ICD extractions from diagnosis notes, alternate VOC recommendations, etc.
- Worked on a framework to evaluate and measure model fairness and bias metrics of our models towards different demographic groups.

### Software Dev Engineer II (ML Platform)

SWIGGY

Jan. 2019 – Jul. 2021  
*DATA SCIENCE PLATFORM, TIME SERIES FORECASTING*

- Was part of team that worked on Feature Store and pipeline which feeds on-demand features to deployed ML models at production scale(4Bn rows, 10K QPS). Pipeline-supported multichannel ingestion, i.e. Spark, Flink and user files etc.
- Founding member of Forecasting and Correlation Platform which was considered by many teams to forecast concerned time series. These forecasts power critical scaling decisions across organizations in real time.
- Led DAQ, a tool used to scrape APIs at scale. Used to collect data for analysis/ model training at a scale of 15 M rows daily.

### Software Development Engineer (Search Relevance)

FLIPKART (a Walmart company)

Sep. 2017 – Jan. 2019  
*SEARCH RELEVANCE, QUERY INTENT, NLP*

- Was responsible for improvements/inception of search intent models(CRF/Neural Network based), identifying error classes, coming up with solutions, and fixing them. These models power user search and discovery for millions every day.
- Implemented a FastText based query store classifier, which predicts the category of a tail query.
- Implemented the first workflow to automate training and auto-deployment of various search models in Flipkart. First was written using Luigi and later migrated to Airflow.
- Wrote a generic framework using Airflow which at runtime creates generic dags for different ML models and orchestrates their training to deployment flow, including data and model validations.
- Implemented large scale (4Bn+ datapoints) pipelines using Cascading/HDFS to extract data from user events and then transform it to be used for training these models.

### Software Development Engineer

GROUPON

Sep. 2016 – Sep. 2017  
*BACKEND ENGINEERING*

- Worked on a component called Cyclops, an interface between Customer representatives and internal services.
- This service is live in all countries in which Groupon operates.

### Software Engineer

NETSPEED SYSTEMS (Acquired by Intel)

Sep. 2015 – Aug. 2016  
*GRAPH ALGORITHMS, NETWORK ON CHIP*

- Led engineering efforts on modules like Polarity based Arbitration, Multi-Cast Filtering, Structural Latency Breakdown, etc.

## PUBLICATIONS

---

Accepted at [NetSci 2023](#) (Poster Presentation) and [IC2S2 2023](#) (Parallel Talk) as **first author**

- [1] **Ashutosh Tiwari**, Prof. Sadamori Kojaku, Prof. Yong-Yeol Ahn, "Biased Contrastive Learning debiases Graph Neural Networks," *International Conference on Network Science (NetSci)*, 2023. [Poster](#)

In this work, we propose a non-parametric contrastive learning framework to learn debiased graph embeddings with respect to sensitive node attributes and structural homophily. Through empirical evaluations on different datasets, we demonstrate that our method offers a better approach to debiasing compared to existing approaches and thus results in more organic recommendations across different GNN architectures.

## EDUCATION

---

### Indiana University, Bloomington

*Master of Science in Computational Data Science 3.87/4.0*

Bloomington, IN

### National Institute of Technology

*Bachelor of Science in Computer Science 8.32/10.0*

Patna, India

## RESEARCH EXPERIENCE / INDEPENDENT STUDY

---

- Implementing novel training frameworks that result in unbiased (or fair) recommendation systems from "biased" datasets at [IUNI](#) with Prof. YY Ahn and Prof. S Kojaku as part of my Independent Study. Summer 2022 - Summer 2023
- Working as a paid RA for, "**User Intent as a Network**" with Prof. YY Ahn, P Kantak and FB Yara. Project is funded by Kelly Business School. Goal is to be able to quantify and thus act upon that "intent" network. Summer 2022 - Fall 2022
- Was part of [NLP Lab](#)@IUB with Prof. D Cavar. Contributed to design of [TieML](#) and Events' Timeline modelling using different fine-tuned Large Language Models (LLMs). Fall 2021

## TEACHING EXPERIENCE

---

Teaching Assistant for [Network Science](#) (INFO-I 606) with Prof. YY Ahn

Spring 2023

Teaching Assistant for [Machine Learning](#) (CSCI-B 555) with Prof. R Khardon

Fall 2022

Teaching Assistant for [Network Science](#) (INFO-I 606) with Prof. YY Ahn

Spring 2022

## SELECTED PROJECTS

---

### BiasNet | Deep Reinforcement Learning, PyTorch, Actor Critic Algorithm, On-policy Model Free

[Code/Report](#)

- Learning to fight in [Street Fighter II](#) with induced relational bias from differential game scenes.

### DeepFoodie | Python, Tensorflow, Self Supervised Deep Clustering, Deep Learning, Transfer Learning

[Code/Report](#)

- Clustering dishes on basis of their ingredient embeddings. These ingredient embeddings are generated by a NN.

### BlindNet | Python, Pytorch, Deep Learning, Transfer Learning

[Code/Report](#)

- Image to vector generation on Coco dataset.

### Continuous Dominant Set Repair | C++, Graph Algorithms, Guha and Khuller's Algo., Greedy

[Code](#)

- Repairs a broken link in Continuous Dominant Set in  $O(\Delta^2)$ , where  $\Delta$  being the avg cardinality of connected graph.

### Humana Mays Healthcare Analytics Case Competition | Boosted Trees, Feature Engineering

[Code](#)

- 11th [rank](#) on leaderboard. Hosted by TAMU and Humana Mays, 2021.

### AnalyticsVidhya Solutions | Python, Tensorflow, Torch, Time Series, Feature Engineering, Catboost

[Code](#)

### Investigating Bias Manifolds | Bias Manifolds, Bias Progression, Measuring Bias, Python, Word2vec

[Code/Report](#)

## TECHNICAL SKILLS

---

Search Relevance, Recommendations, Graph Neural Networks, Fairness Aware Modeling, Computational ML, NLP, Computer Vision, DL, Spark, Python, Scala, C++, Contrastive Learning, Large Language Models

**Frameworks/Libraries/Tools:** Ray, Pytorch Geometric, Pytorch Lightening, Pytorch, Tensorflow, Sklearn, Numpy, Pandas, Matplotlib, HDFS, Spark, Kafka, Flink, AWS Sagemaker, DynamoDB, Faiss, vLLM, Ray, Django