

Aman Singhal

551-344-6364 | as14661@nyu.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

New York University

Masters of Science in Computer Science

Courses: Deep Learning, Probability and Statistics, Data Structures and Algorithms

New York, NY

August, 2022 - May, 2024

Delhi Technological University

Bachelors of Technology in Electronics and Communications Engineering

Courses: Computer Vision, Computer Architecture, Calculus, Linear Algebra

Delhi, India

August, 2014 - May, 2018

EXPERIENCE

UnitedHealth Group

Data Scientist [Awards](#)

Delhi, India

July, 2021 - July, 2022

- Implemented Computer Vision and Natural Language Processing tasks, enabling intelligence across medical/legal documents with **limited domain data**.
- Won global **hackathon** with Transformer based Rx information extraction; developed product, '21 revenue \$ 9M.
- Fine-tuned PyTorch T5 & BART encoder decoder for legal **text generation**; backtranslated data, ROUGE 85%.
- Led **BERT** named entity recognition & RNN text classification; preempted breach of data privacy, saving \$20M.
- Automated drug guideline recommendations; clustered USE Encoder via dimensionality reduction, timecost 30%.

CVIT Lab, IIIT

Research Assistant (Prof C V Jawahar) [Paper](#) [Code](#)

Hyderabad, India

August, 2020 - March, 2021

- Exploring Pairwise NMT for Indian Languages **published** at International Conference on NLP, ACL Anthology.
- Pioneered computationally inexpensive **transfer learning** technique for low-resource **domain adaptation** and neural machine translation; SOTA +18.0 BLEU.
- Achieved leaderboard across 10 language modeling tasks at the **Workshop** for Asian Translation (WAT'20).

TransOrg Analytics

Data Scientist [Paper](#) [Code](#)

Gurugram, India

June, 2018 - August, 2020

- Correlated word2vec performance on word analogy tasks and sentiment analysis, sped up LSTM training by 50%.
- Scraped** 5B merchant reviews for American Express with Beautiful Soup; pre-processed social multilingual text.
- Ensembled **TF-IDF** Random Forest, XGBoost & SVM, and GloVe **RNN**; classified 0.1% event-rate 95% F1.
- Led insight product at AmEx with statistical modeling & exploratory analysis; generated \$3M via UX analytics.

RESEARCH PROJECTS

Semantic Search Engine, Contract Intelligence [Code](#)

- Developed YOLO object detection, segmentation and Elasticsearch + BERT pipeline; ranked 0.1B clauses in 90s.
- Deployed **backend API** with Docker & Kubernetes; optimized NoSQL **database design** for real-time search.

Audio Analytics, Content Moderation (Prof Aaditeshwar Seth, IIT Delhi) [Code](#)

- Automated spectral CNN & ResNet speech classification; developed **AWS** pipelines for non-profit social platform.
- Catalyzed time-cost saving with CatBoost, LightGBM and frequency domain features, scaled 1M rural household.

Insurance Risk, Predictive Modeling (Montana State)

- Setup Hadoop **ETL** & **Spark** & Hive wrangling pipelines; led regression, classification, & time-series forecasting.
- Collaborated with operations team to develop interpretable Tableau dashboards, claims reduced 7% YOY.

Low-cost Human Feedback (Prof Vinay Namboodiri, Univ of Bath, UK) [Report](#)

- Researched active learning & reinforcement learning; regulated self & weak supervision; lowered annotation cost.
- Developed uncertainty measures; tuned hyperparameters & cross-validated; maintained cross-entropy loss curve.

SKILLS

Machine Learning: Deep learning (Keras, PyTorch, TensorFlow), NLP (HuggingFace, spaCy), Scikit-Learn, OpenCV
Languages: [Python](#), SQL, R, C++, C, JavaScript, Scala, Go, React, SAS, MATLAB, Bash
Software: Jupyter, APACHE, Large-scale data processing, PySpark, Hive, MySQL, Microsoft Excel, Power BI, \LaTeX
Production Systems: Kubernetes, Cloud platforms (AWS EC2, S3), Docker, Git, Flask, TF Serving, CI/CD (Jenkins)
Data Science: Numpy, Pandas, A/B Testing, Data Visualization (Tableau, Matplotlib), Time series (Statsmodel), SciPy