Aman Singhal

linkedin.com/in/amansinghalml | +1 (551)-344-6364 | github.com/AmanSinghal927| as14661@nyu.edu | https://amansinghal927.github.io/

EDUCATION

New York University — New York, NY
Master of Science, Computer Science — Sep 2022 - Present
NLP, Large Language Models (LLMs), Artificial Intelligence, Machine Learning, Algorithms, Data Structures, Optimization

Delhi Technological University — Delhi, India
Bachelor of Technology, Electronics and Communication Engineering — Aug 2014 - May 2018
Computer Vision, Mathematics (Calculus, Linear Algebra, Probability, Statistics), Signal Processing, Economics

EXPERIENCE

Data Science Intern, Chegg Inc., CA, USA — May 2023 – Aug 2023
- Devised mathematical reasoning prompts for Physics, boosting gpt-4 automated evaluation metrics by 2x over the baseline
- Enhanced helpfulness of inhouse LLM for Biology dataset by 9% using reinforcement learning from human feedback (RLHF)
- Spearheaded retrieval augmented generation POC, improving the factual consistency of generative neural networks

Data Scientist, UnitedHealth Group, Delhi, India | *Awards* — July 2021 – July 2022
- Applied information extraction research, finetuning legal + healthcare computer vision & nlp models to pharma documents
- Led agile teams, streamlining workflow automation & cloud deployment of multi-modal models for prior-authorization
- Built image classification, image segmentation & object detection pipeline to ingest tables from UHG's contracts in real-time
- Drove search engine, RNN ner & huggingface transformers question answering web platform to query 1B pbm clauses in 9s

Research Assistant, IIIT Hyderabad (Prof. Vinay Namboodiri & Prof. C V Jawahar) | *Papers* | *Code* — Aug 2020 – Mar 2021
- Co-authored computationally inexpensive transfer learning research for low-resource domain adaptation and neural machine translation using PyTorch transformers & AWS. Achieved state-of-the-art +18.0 BLEU over existing deep learning algorithm
- Open-sourced semi-supervised synthetically generated dataset for language translation, advancing generative AI research

Data Scientist, TransOrg Analytics, Gurugram, India | *Papers* | *Code* — June 2018 – Aug 2020
- Managed cross functional team & led requirement gathering, communicating business intelligence to client leadership
- Automated loan underwriting, unsupervised fraud prevention and economics models for insurance risk at American Express
- Trained resource allocation regression models using exploratory analytics (eda) & big data mining for the Montana state govt.
- Led feature engineering at Nissan, collaborating with marketing to AB-test mathematical models for customizing campaigns

SKILLS

Machine Learning: Numpy, Pandas, Scikit-Learn, data vizualization (Tableau, Matplotlib), Statistical Learning (SciPy)
Deep learning: NLP (OpenAI, LangChain, HuggingFace, spaCy), AI/ML (PyTorch, Keras, TensorFlow), WandB
Languages: Python, C++, SQL/ MySQL, R, scala, apache spark, Hive, latex, ui/ux (javascript), business intelligence (tableau)
Software: Kubernetes, Cloud (aws, azure), Docker, Git, PineCone, Kubeflow, trlx, TF Serving, Jupyter, Apache, ETL (hadoop)

PROJECTS

LLM Generation, Chegg: Explored iterative refinement, multiple personas & chain of thoughts prompting 1.5x stepwise quality
- Quality aligned gpt-4 using proximal policy optimization (PPO), deepspeed & parameter efficient finetuning (PEFT) LoRA
- Developed annotation & reward finetuning for generative AI, achieving 75% accuracy with Flan-T5 & DeBERTa models

Semantic Search (code): Transformed legal research via azure ocr, pretrained bert & yolo object detection multi-modal pipeline
- Streamlined backend Python APIs for parse tree summarization with Docker, Kubernetes & NoSQL database management
- Supported javascript UX & managed ETL, tf serving pca dimensionality reduction, k-means clustering & RNN topic model

Audio Analytics, IIT Delhi (code): Led signal processing, multi-modal feature engineering & data vizualization for social platform
- Automated xgboost, CNN & ResNet speech classification; developed AWS pipelines for real-time mobile content moderation
- Catalyzed time-cost saving with hyperparameter tuning, data augmentation & feature engineering, scaled 1M household

Risk Modeling, Montana State: ETL big data through Hadoop & Spark for predicting insurance claims with 80% F1
- Managed 1% event-rate using random forest & gradient boosting cascade, weighing and sampling to reduce accidents by 7%

Demand Forecasting, Nissan: Implemented end to end hive data science pipeline, predicting servicing churn with 87% accuracy
- Handled preprocessing, CI/CD for SVM & neural network ensemble & AB-testing, increasing customer retention by 15%

Video frame prediction & collision modelling, NYU: Leveraged signal processing & self-supervision with SimVP & U-Net for future segmentation prediction in video dataset, won 1st place in Prof. Yann LeCun's Deep Learning class

PUBLICATIONS & OPEN-SOURCE CONTRIBUTIONS

- **Published** Exploring Pairwise NMT for Indian Languages research at the ICON conference, ACL Anthology link
- Developed leaderboard neural deep learning transformer models at the **workshop** for language translation (WAT'20 - link)
- Won UHG's hackathon by developing statistical AI algorithm for medical information extraction using JavaScript & RNNs
- Achieved excellent grade in post graduate computer science degree, learning full-stack dev & object-oriented programming
- Teaching assistant for applied mathematics & artificial intelligence courses in the computer science department at NYU
- Trained 7 junior developers in Python, hive, data visualization, business intelligence & AB-testing at TransOrg Analytics