

# Ashish Gupta

ashug219@gmail.com • +91 9088467250

Web - ashish-gupta03.github.io

LinkedIn: ashishgupta031

INTERESTS	Machine Learning, Natural Language Processing, Information Extraction, Information Retrieval, Recommender Systems, Reinforcement Learning, Generative AI
EDUCATION	<p><b>M.Tech(CSE) with specialization in Machine Learning</b></p> <p>▪ <b>IIIT Bangalore</b> Jul 2016 – Jul 2018</p> <ul style="list-style-type: none"><li>• Thesis: Neural Attention Reader for Video Comprehension</li><li>• Adviser: Prof. Manish Gupta</li><li>• Focus: Information Retrieval, Information Extraction, Videos, Attention Mechanism, Bidirectional LSTM, Distant Supervision, Differential Weighing</li></ul> <p><b>B.Tech in CSE</b></p> <p>▪ <b>SRMCEM(Uttar Pradesh Technical University), Lucknow</b> Aug 2008 – Jun 2012</p> <ul style="list-style-type: none"><li>• Deans List for 3 years.</li></ul>
RELEVANT EXPERIENCE	<p><b>ServiceNow</b></p> <p>▪ <b>Staff ML Engineer</b> Dec 2023 – Present</p> <ul style="list-style-type: none"><li>• Worked on automated code generation, explanation, and completion.</li><li>• Working on question-answering and multilingual retrieval systems (embedding + reranking + LLM). Fine-tuned and evaluated SOTA embedding models to see which works best for our use case.</li><li>• Working on multimodal RAG which involved texts, images, and unstructured text, and finally evaluated via LLMs.</li></ul> <p><b>Sharechat</b></p> <p>▪ <b>Lead ML Engineer</b> Aug 2022 – Sep 2023</p> <ul style="list-style-type: none"><li>• Working in recommendations, personalization, and ranking domains.</li><li>• Working on in-session personalization and ranking stack mainly involves creating DAG via Airflow, fetching data from BQ, and using GCS to store and retrieve data.</li><li>• Used TF records and built ML models around the TensorFlow stack via distributed training on TPUs.</li><li>• Evaluate the model on multiple metrics offline. Add Slack and Opsgenie alerts. Run A/B Tests to evaluate online.</li></ul> <p><b>Microsoft</b></p> <p>▪ <b>Applied Scientist II</b> Apr 2021 – Jul 2022</p> <ul style="list-style-type: none"><li>• Worked in the Defensive Ranker team where we needed to rank URLs/documents based on the nature of the threat.</li><li>• Ran Interleaving flights, performed A/B testing and used multiple metrics to measure the performance.</li><li>• Developed multi-lingual neural models for the Defensive Search team where we work on specific threats like coronavirus, black lives matter, anti-semitic, suicide, etc. to suppress URLs based on their nature.</li><li>• Worked in the creation of a Universal metric measurement set and used multiple Bayesian sampling methods for scaling across different languages and regions</li></ul> <p><b>Walmart Global Tech</b></p> <p>▪ <b>Data Scientist</b> Aug 2019 – Apr 2021</p> <ul style="list-style-type: none"><li>• Worked in the Catalog Data Science team.</li><li>• Built deep learning models for attribute extraction from text.</li><li>• Developed a Smart normalization tool to match the non-standard/junk text present in the catalog to the standard text.</li><li>• Developed BERT-based models for classification and sequence labeling in Multi-lingual models. These models are further used in product search.</li><li>• Led an initiative to build an auto tagger tool for reducing the amount of tagged data and building efficient models with limited data.</li><li>• Lead an initiative on discovering high-quality entities/attributes from Walmart product reviews. This will help to enrich the catalog in a more fine-grained manner.</li></ul> <p>▪ <b>Senior Statistical Analyst</b> Jul 2018 – Jul 2019</p> <ul style="list-style-type: none"><li>• Worked on retail graph for the home and furniture section, which includes entity extraction, and style prediction. Built models that helped in substitute and complementary product recommendations.</li></ul> <p><b>VideoKen Software Pvt. Ltd, IIIT Bangalore Innovation Centre</b></p> <p>▪ <b>Visiting Researcher</b> Jan 2018 – Jun 2018</p> <ul style="list-style-type: none"><li>• Worked on the neural multi-task reader for Video Comprehension. Used Attention Mechanism, Bidirectional LSTM, and self-attention and did meaningful bifurcations of the raw text to complete the task.</li></ul>

## Tata Consultancy Services

### ▪ Data(Systems) Engineer

Nov 2012 – Jun 2016

- Worked in Oracle apps (an ERP tool) as an OTR consultant. Worked on GE Healthcare projects.

## PUBLICATIONS

### ▪ Defensive Low Authority Host Predictor

- Ashish Gupta, Sunakshi Gupta, Somi Satti Reddy in Microsoft Journal of Applied Research(MSJAR) 2022.

### ▪ Learning with Limited Labels via Momentum Damped Differentially Weighted Training

- Rishabh Mehrotra, Ashish Gupta in KDD 2020.

### ▪ Joint Attention Neural Model for Demand Prediction in Online Marketplaces

- Ashish Gupta, Rishabh Mehrotra in NLDL 2020.

### ▪ Hyper-parameter optimization with REINFORCE and Masked Attention Auto-regressive Density Estimators

- Chepuri Shri Krishna, Ashish Gupta, Swarnim Narayan, Himanshu Rai, and Diksha Manchanda got accepted in IEEE BigData 2020.

### ▪ Ultron-AutoML: an open-source, distributed, scalable framework for efficient hyper-parameter optimization

- Swarnim Narayan, Chepuri Krishna, Varun Mishra, Abhinav Rai, Himanshu Rai, Chandrakant Bharti, Gursirat Singh, Ashish Gupta, and Nitinbalaji Singh in IEEE BigData 2020.

### ▪ Sequence-aware Reinforcement Learning over Knowledge Graphs

- Ashish Gupta, Rishabh Mehrotra in RecSys REVEAL 2019.

### ▪ Neural Attention Reader for Video Comprehension

- Ashish Gupta, Rishabh Mehrotra, Manish Gupta in KDD Deep Learning Day 2018.

## PATENTS

### ▪ Ultron-AutoMLv2: a distributed framework for efficient hyper-parameter optimization (HPO) of ML models

- Chepurishri Krishna, Amit Agarwal, Ashish Gupta, Swarnim Narayan, Himanshu Rai, Varun Mishra, Abhinav Rai, Chandrakant Bharti, Gursirat Singh and Nitinraj Balajisingh

## BLOGS

### ▪ An Introduction to Meta-Learning

### ▪ Introduction to Reinforcement Learning

## ACHIEVEMENTS / CO-CURRICULAR ACTIVITIES

- Placed 506 out of 2662 teams in **LLM Science Exam** - Use LLMs to answer difficult science questions. (Used instruction-tuned Llama model)
- Top 3%(Placed 94 out of 4037 teams) in **Quora Insincere Question Classification** - To identify and flag insincere questions in Quora.
- Top 1.4%(Placed 28 out of 2000 teams) in Microsoft AI India Challenge 2018 - Ranking passage according to relevance containing answers to a given question.
- Achieved AIR 56 in ISRO Scientist/SC exam(July'16).
- Qualified GATE'16 with 98.8 percentile(Feb'16).

## PROFESSIONAL AFFILIATIONS & ACTIVITIES

Reviewer of CIKM: Conference on Information and Knowledge Management(CIKM 2024)	2024
Reviewer of Empirical Methods in Natural Language Processing(EMNLP 2023)	2023
Reviewer of Knowledge Discovery and Data Mining(KDD 2021)	2021
Reviewer of Association for Computational Linguistics(ACL 2020)	2020
SIGIR Conference on Research and Development in Information Retrieval(SIGIR 2019)	2019
Association for Computing Machinery	2017 – Present

## SKILLS

- **AI/ML:** Tensorflow, PyTorch, Keras, NLTK, scikit-learn, spaCy
- **Optimization:** CVXPY
- **Programming Languages:** C, Java, Python
- **Database:** Hive, MySQL, MongoDB, MS SQL Server, BigQuery
- **Deployment:** Airflow, Docker, Tensorflow serving, Google cloud