

GAURAV GUPTA

Amazon, Applied Scientist II, Palo Alto, CA

LinkedIn: <https://www.linkedin.com/in/gauravgupta16/>

Email- gauravgupta16may93@gmail.com

Github- <https://github.com/gaurav16gupta>

PROFESSIONAL SUMMARY

Applied Scientist with 5+ years of experience in designing and optimizing machine learning models for large-scale information retrieval and recommendation systems. Proficient in Python, C++, and PyTorch. Published researcher with a proven track record of improving business outcomes through innovative solutions.

SKILLS

- Technical Skills: Python, Numpy, Pandas, Spark, C++, PyTorch, TensorFlow
- Machine Learning: Recommendation Engines, Large Language Models(LLM), Search, Ranking, NLP, Generative AI, Retrieval Augmented Generation (RAG), Deep Learning, Computer Vision, Computer Science
- Soft Skills: Research and Documentation, Communication (verbal and written), Mentorship

PROFESSIONAL EXPERIENCE

Amazon, Applied Scientist II, *Palo Alto, CA, Nov 2023 - Present*

- Designed and implemented foundational models for personalized search. Used customer behavior data, and optimized production ANN index to enhance user engagement and sales.
- Formulated query reformulation strategies with LLMs, deployed and conducted production A/B testing, found significant revenue growth within key markets.
- Improved semantic search models by fusion training with open source LLM embeddings.
- Collaborated with cross-functional teams. Published and presented at an internal conference.

Adobe Research, Research Intern

June 2022 - Aug 2022

- Developed data drift detection techniques for recommendation models

June 2021 - Aug 2021

- Proposed a product recommendation algorithm with dense-semantic and sparse-lexical features while beating the state-of-the-art methods on retrieval time. (Work published in Big Data, 2023)

Amazon, Applied Scientist Intern

June 2020 - Aug 2020

- Proposed a billion-scale near-neighbor index, optimizing load balance and computational efficiency (work published at SIGKDD 2022)

June 2019 - Aug 2019

- Worked on efficient sublinear and low-memory indexing methods for constraint filtering

TCS Innovation Labs, Researcher (R&D)

Delhi, India, July 2014 - Aug 2018

- Developed machine learning-based solutions for retail, augmented reality devices, and indoor localization. Published five papers in top-tier conferences

EDUCATION

Rice University, Houston,

Ph.D. in Large-Scale Information Retrieval ([Thesis](#)), Houston, TX, Aug 2018 - Aug 2023

- Thesis Advisor: Prof. Anshumali Shrivastava
- Coursework: Probabilistic Data Structures, Machine Learning, NLP, Optimization, Multi-Core Programming

Indian Institute of Technology, Hyderabad

B.Tech in Electrical Engineering, Hyderabad, India, Aug 2010 - May 2014

PUBLICATIONS

- Identity with Locality: An ideal hash for efficient gene sequence search, KDD 2025, *Gaurav Gupta**, *Tianyi Zhang**, *Aditya Desai**, *Anshumali Shrivastava*
- CAPS: A Practical Partition Index for Filtered Similarity Search, ISIR-eCom, WSDM 2024, *Gaurav Gupta*, *Jonah Yi*, *Benjamin Coleman*, *Chen Luo*, *Vihan Lakshman*, *Anshumali Shrivastava*
- Near Neighbour Search for Constraint Queries , IEEE International Conference on Big Data, 2023, *Gaurav Gupta*, *Anup Rao*, *Tung Mai*, *Ryan A Rossi*, *Xiang Chen*, *Saayan Mitra*, *Anshumali Shrivastava*
- BLISS: A Billion scale Index using Iterative Re-partitioning, KDD 2022, *Gaurav Gupta*, *Tharun Medini*, *Anshumali Shrivastava*, and *Alexander J. Smola*
Invited talk at NeurIPS 2021 BigANN competition track <https://big-ann-benchmarks.com>
- Fast Processing and Querying of 170TB of Genomics Data via a Repeated And Merged BloOm Filter (RAMBO), 2226-2234, ACM SIGMOD, 2021, *Gaurav Gupta**, *M. Yan**, *B. Coleman*, *R. A. Elworth*, *T. Medini*, *T. Treangen*, *A. Shrivastava*
Featured News Article: RAMBO speeds searches on huge DNA databases, <https://www.sciencedaily.com/releases/2021/06/210628152920.htm>
- Siamese Networks for chromosomes classification, BIC, ICCV 2017 (pp. 72-81), *Swati*, *Gaurav Gupta*, *Mohit Yadav*, *Monika Sharma*, *Lovekesh Vig*
- Indoor Localisation and Navigation on Augmented Reality Devices, IEEE ISMAR, 2016, *Gaurav Gupta*, *N. Kejriwal*, *P. Pallav*, *E. Hassan*, *S. Kumar*, *R. Hebbalaguppe*

Refer additional publications at Google Scholar: <https://scholar.google.co.in/citations?user=OguKfJIAAAAJ>

PATENTS

- Sparse embedding index for search |US Patent US20230153338A1| *Gaurav Gupta*, *Anup Rao*, *Tung Mai*, *Ryan A. Rossi*, *Xiang Chen*, *Saayan Mitra*
- Method and system for extracting information from hand-marked industrial inspection sheets | EP Patent 201721039681 | *Gaurav Gupta*, *Swati Jindal*, *Monika Sharma*, *Lovekesh Vig*
- System and methods for detecting and tracking a marker, US Patent 20,170,278,266 | *Gaurav Gupta*, *Ehtesham Hassan*

ACHIEVEMENTS

- Invited talk: How to Deal with Volume and Velocity Associated with Hundreds of Terabytes (and Beyond) of Genomics Data? - 2021 Ken Kennedy AI and Data Science Conference.
- Travel Grants - MLSys 2022 (NSF Funded), NeurIPS 2019.
- Ph.D. panellist at undergrad student mentorship program at KDD 2022.
- Nominated by the ECE department at Rice University for Google Ph.D. fellowship.
- Ranked first at Nielsen's Image Recognition Hackathon- 2017 for "Hole detection/ product out of stock" retail store challenge.
- Ranked 1913 (among top 0.4 %) in IIT-JEE engineering entrance exam, 2010.
- Selected for Deep Learning Summer School- MILA, Montreal, 2016.

ACTIVITIES AND LEADERSHIP

- Mentor for undergraduate interns at Rice University (Summer 2022)
- Academic Mentor for Rice ECE Graduate Student Association (2020)
- Reviewer for top-tier conferences: AISTAT-2021, CIKM-2022, Neurips-2022, ICLR-2022, ICML-2023, KDD-2023, Neurips-2023, KDD-2024
- Teaching assistant- Introduction to Deep Learning, Probabilistic Algorithms and Data-structures (2020)
- Academic Mentor for Rice ECE Graduate Student Association (2020)
- Media Head (photography and videography) at Indian Students at Rice (ISAR) (2019-20)
- Student Coordinator at Training and Placement Cell- IIT Hyderabad. (2013-14)
- Electronics Club Coordinator- IIT Hyderabad (2012-13)