

Anshumaan Chauhan

achauhan@umass.edu | 321-361-9962 | [linkedin.com/in/anshumaan-chauhan/](https://www.linkedin.com/in/anshumaan-chauhan/) | github.com/Anshumaan-Chauhan02

EDUCATION

University of Massachusetts Amherst

Master of Science (MS) in Computer Sciences

Sep 2022 – May 2024

GPA- 3.93/4

Birla Institute of Technology and Science Dubai

Bachelor of Technology (BTech) in Computer Sciences

Aug 2018 – June 2022

GPA- 9.83/10

Relevant Coursework: Algorithms for Data Science, Systems for Data Science, Machine Learning, Artificial Intelligence, Natural Language Processing, Reinforcement Learning, Software Engineering, Neural Networks, Object Oriented Programming, Database Management Systems, Data Structures and Algorithms.

SKILLS

Programming Languages: Python | Java | HTML | CSS | TypeScript | SQL | NoSQL | LaTeX

Amazon Web Services: EC2 | Lambda | Batch | S3 | DynamoDB | Athena | Step Functions | SNS | SES | SDK | CDK

Frameworks/Tools: Tableau | SLURM | Git | Postman | Docker | Airflow | Snowflake | Apache Spark | Linux |

Streamlit | Kubernetes | MySQL | PostgreSQL | MongoDB | DuckDB | JIRA | Confluence | Lucidchart

Machine Learning Tools: PyTorch | Tensorflow | NumPy | pandas | scikit-learn | Matplotlib | Seaborn | Transformers

PROFESSIONAL EXPERIENCE

Data Engineer, Bose Corporation, Framingham, MA

July 2024 - Present

- Designed and implemented secure event-driven ETL pipelines using AWS serverless components (Lambda, Step Functions, Secrets Manager), achieving a 5% cost reduction through efficient secret consolidation and provisioning automation.
- Utilized SQL transformations and Snowpark for seamless Python integration to optimize data extraction processes for PowerBI Cloud, enhancing data loading into Snowflake and reducing operational costs by 10%.
- Created alert and failure monitoring pipelines with a Streamlit dashboard for Snowflake-deployed apps, leveraging telemetry logs and reducing downtime by 25% through daily queries via cron jobs.

Graduate Student Researcher, Amazon, Amherst, MA

Feb 2024 – Jun 2024

- Conducted experiments to enhance small language models (SLMs) with custom reasoning chains, identifying limitations in zero-shot performance.
- Proposed a decoupled three-step process (plan, reason, answer extraction) that improved GSM8K benchmark scores by +0.57 for 3B models and +3.32 for 7B models.
- Developed and tested a Partial Self Consistency (PSC) method for ensemble analysis, demonstrating that PSC outperforms standard self-consistency by up to +1.85 across various reasoning techniques.

Machine Learning Researcher, Florida Institute of Technology, Melbourne, FL

Jan 2022 – Jun 2022

- Analyzed and extracted the representation of the specifications in a subset of English language using Natural Language Processing (NLTK library) and designed a compiler for translating it to AADL. [\[Paper\]](#)
- Proposed a Double Deep Q-Network for NAS, reducing scalability and time complexity while maintaining 97% of SOTA performance with ~35% fewer layers. [\[Paper\]](#)
- Developed an unsupervised DBSCAN-based approach for medical image segmentation, achieving Dice Coefficients of 0.784 and 0.88 for kidneys and lungs, while avoiding the need for large, labeled datasets. [\[Paper\]](#)

Student Researcher, Birla Institute of Technology and Science, Dubai, United Arab Emirates

Aug 2020 – Dec 2021

- Developed a denoising autoencoder architecture with Z-score normalization, Robust PCA for data denoising, and projected gradient descent, resulting in a 95.8% accuracy in novelty detection, outperforming state-of-the-art models by 2.2-4.3%. [\[Paper\]](#)
- Enhanced Support Vector Machines with a genetic algorithm inspired by Reinforcement Learning, boosting accuracy from 6.12% to 36.52%, and achieved a 3rd generation model accuracy of 93.88%. [\[Paper\]](#)

PROJECTS

[Guided Conditional Image Generation with Conditional Flow Matching](#) | Python, PyTorch, NumPy, pandas

- Integrated Conditional Optimal Transport into an attention-based UNet model for image generation, achieving FID scores of 105.54 (unconditional) and CLIP/FID scores of 22.19/385.56 (conditional).
- Utilized the BLIP2 FLAN T5 model for image captioning, addressing descriptive limitations of the CIFAR10 dataset and improving image-text alignment for better interpretability.

[Recipe Infusion](#) | Python, Transformers, Matplotlib, Sklearn

- Created a Recipe Infusion framework featuring Recipe Generation and Style Transfer components, enhancing culinary creativity and personalization.
- Fine-tuned the DistilGPT model on combined RecipeNLG and RecipeBox datasets, improving BLEU and Perplexity scores by +0.04 and 24.5 points, respectively, and successfully implemented style transfer for various celebrities using T5-small models.