Anshumaan Chauhan

achauhan@umass.edu | 321-361-9962 | linkedin.com/in/anshumaan-chauhan/ | github.com/Anshumaan-Chauhan02

Education

University of Massachusetts Amherst, Amherst, MA **2022 – 2024**

Master of Science (MS) in Computer Sciences *GPA- 3.93/4*

- Relevant Coursework: Algorithms for Data Science, Systems for Data Science, Machine Learning, Artificial Intelligence, Natural Language Processing, Reinforcement Learning, Software Engineering, Neural Networks.

Professional Experience

Bose Corporation, Framingham, MA - *Data Engineer CoOp* **2024**

- Developed secure and scalable AWS pipelines using Lambda, Secrets Manager, and Step Functions, reducing costs by 5% through secret consolidation and automating provisioning with AWS CDK.
- Optimized PowerBI Cloud data extraction by minimizing AWS Step Functions state transitions, cutting operational costs by 10% and streamlining Snowflake data loading.
- Built a Streamlit app deployed on Snowflake for efficient JIRA ticket creation and utilizing AWS pipelines to automatically fetch relevant data from JIRA and create new tickets.
- Researched the benefits of migrating from Hive to Iceberg tables and explored Snowflake best practices and query optimization to effectively reduce overall Snowflake costs.

Amazon, Amherst, MA – *Graduate Student Researcher* **2024**

- Conducted experiments to enhance reasoning capabilities of small language models (SLMs) using custom reasoning chains, highlighting limitations in zero-shot performance and proposing a decoupled three-step process (plan, reason, answer extraction) which improved GSM8K benchmark scores by +0.57 (3B models) and +3.32 (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.

Florida Institute of Technology, Melbourne, FL – *Machine Learning Researcher* **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.
- Proposed an approach using Double Deep Q Networks for the automated generation of Convolutional Neural Network architectures, minimizing the scalability and time complexity problems without having effect on Search Space by implementing One Shot Training and Prioritized Experience Replay.

Projects

Guided Conditional Image Generation with Conditional Flow Matching (Python, PyTorch, NumPy, Pandas) **2024**

- Developed a unified attention-based UNet model incorporating Conditional Optimal Transport and Classifier Free Guidance (CFG) for proficient conditional and unconditional image generation, achieving FID scores of 105.54 and CLIPScore/FID scores of 22.19/385.56.
- Enhanced image captioning for the CIFAR10 dataset using the BLIP2 FLAN T5 model, addressing descriptive limitations and significantly improving the quality of generated captions.

Recipe Infusion **2023**

- 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.

Scalability Check for Machine Learning System Predicting Flight Delays **2023**

- Developed a robust pipeline leveraging MySQL for data storage and Spark for querying, ensuring efficient flight delay predictions.
- Evaluated scalability by measuring MySQL and SparkSQL response times across varying dataset sizes, achieving linear latency growth and successful feature extraction and ML algorithm application for accurate predictions.

Technical Skills

Programming Languages: Python | SQL | Java | HTML | CSS | TypeScript | Lex | Yacc
Databases: MySQL | MongoDB | DuckDB
AWS - EC2 | Lambda | Batch | S3 | DynamoDB | Athena | Step Functions | SNS | SES | SDK | CDK
Applications/Tools: Tableau | Git | Postman | Docker | Airflow | Snowflake | PySpark | Linux | Streamlit
Productivity Software: JIRA | Confluence | Lucidchart