Anugrah Vaishnav

Lowell, MA (978)-569-6430 vaishnav.anugrah123@gmail.com linkedin.com/in/anvaishnav anugrahvaishnav.github.io

EDUCATION

M.S. in Computer Science Sep 2023 - May 2025
University of Massachusetts Lowell **(GPA - 3.85/4)** *Lowell, MA*

B.E. in Electronics Sep 2017 - May 2021
Ramaiah Institute of Technology, India **(GPA - 8.3/10)** *India*

WORK EXPERIENCE

Financial Data Analyst | *University of Massachusetts Lowell* [*Topic Modeling, NLP*] Mar 2025 - May 2025

- Extracted key topics from IPO financial reports to improve the interpretability of the report for stakeholders.
- Implemented Latent Dirichlet Allocation (LDA), and TF-IDF vectorizer to enhance financial text analysis and identify critical disclosure themes.

Data Analyst | *Portcast* [*Time Series Forecasting, SQL, Tableau*] Aug 2022 - May 2023

- Contributed to Port Arrival Forecasting efforts by integrating real-time maritime data (Automatic Identification System), refined feature engineering pipelines, **improved ETA prediction accuracy by 7%**.
- Automated report generation workflows, **saved 10+ hours of manual QA weekly**; built Tableau dashboards to monitor port delays and vessel performance.

Data Analyst | *Equipped Analytical Intelligence* [*Time Series Forecasting, SQL, Tableau*] Nov 2021 - Aug 2022

- Worked on predictive modeling for cash flow forecasting, improving the accuracy of money lending scenario analysis for alternative investors; integrated granular portfolio data into the company's SaaS intelligence platform.
- Migrated the visualization code base to a modern tech stack in Shiny-R, **increased reporting efficiency by 30%**, and automated multi-stakeholder reports, **saved 10+ hours of manual QA monthly**.

PROJECTS

Enhanced 3D Reconstruction in Colonoscopy | *Advisor: Yu Cao* [*Monocular Depth Estimation, Computer Vision*]

- Fine-tuned DepthAnything V2 on the SimCol3D synthetic colonoscopy RGB-D dataset; **surpassed prior SOTA by 10–15%** on L1 and RMSE metrics for depth estimation.
- Built an end-to-end pipeline integrating depth and pose predictions to generate dense point clouds and 3D meshes via Ball Pivoting and Poisson reconstruction.
- Leveraged **multi-GPU fine-tuning** with parameter optimization, mixed precision, and gradient accumulation to improve training time and resource efficiency.

Correlated Topic Models [*NLP, Open Source Software*]

- Developed an open-source PyTorch implementation of Correlated Topic Models (CTM) using Automatic Differentiation Variational Inference (ADVI).
- Applied scalable batched training to handle large datasets efficiently.
- Optimized performance, **achieving 10-15% faster processing and 8-10% higher accuracy** compared to libraries such as tomotopy, PyCTM, and topicmodels by integrating ADVI.

Machine Unlearning for Multimodal LLMs [*Machine Unlearning, VLLM*]

- Implemented a **machine unlearning framework** to selectively 'forget' data in LLaVA multimodal LLM.
- Fine-tuned LLaVA using LoRA adapters, implemented optimization and distributed computing with multiple GPU support using DeepSpeed.
- Implemented gradient-based unlearning algorithms; evaluated the model using ROUGE and LLM-as-judge metrics.

Effect of antidepressants on COVID-19 trajectory | *Advisor: Rachel Melamed* [*Causal Inference, Biomedical Informatics*]

- Investigated the causal relationships between the use of antidepressants by patients and their COVID-19 outcomes.
- Implemented causal machine learning models: S/T learners, TARNet, graph neural networks (GNN) and variational autoencoder (VAE) in PyTorch to predict Individual Treatment Effects (ITE) from Electronic Health Records (EHR).
- Identified potential **links between antidepressant use and reduced mortality risk** in patients with COVID-19.

Chest X-ray Report Generation [*Retrieval Modeling, Chest X-ray, CLIP*]

- Developed a retrieval-based system to automatically generate clinical reports for chest X-rays by leveraging a pre-trained CLIP model to map visual data to medical text.
- Fine-tuned CLIP for MIMIC-CXR dataset; retrieve the most semantically similar sentences for input X-ray.
- Evaluated retrieval quality using BLEU scores for text similarity; F1, precision against CheXbert labeler to ensure medical relevance.

SKILLS

- **Languages**: Python, R, SQL (PostgreSQL, MySQL) , NoSQL (MongoDB, DynamoDB), C++, Streamlit
- **Frameworks & Libraries**: PyTorch, Tensorflow, Scikit-learn, Langchain, LangGraph, RAG, Pydantic, Spark (PySpark)
- **Machine Learning**: Deep Learning (CNN, RNN), Generative Modeling, Statistics, Transformers, Computer Vision, NLP, Fine-tuning (LoRA, PEFT, Deepspeed), Large Language/Vision Models (LLM/LVM)
- **Tools**: Docker, AWS, Tableau, Shiny-R