

Aryaman Velampalli

+1 646-549-5845 | akv2129@columbia.edu | linkedin.com/in/aryaman-velampalli | github.com/fin1cky

EDUCATION

Columbia University

M.S. in Computer Science, ML Track; GPA: 4.16

New York, NY

Dec 2026

Birla Institute of Technology and Science, Pilani (BITS Pilani)

B.E. in Computer Science

India

Jul 2023

EXPERIENCE

Graduate Research Assistant

AlQuraishi Lab (AQLab), Columbia University

New York, NY

Feb 2026 – Present

- Conducting research in generative modeling for protein design, with a current focus on diffusion-based methods and reward-guided optimization.
- Building PyTorch prototypes for protein backbone generation and evaluation, studying how guidance and fine-tuning can steer generated structures toward target properties.

Machine Learning Engineer II

Avalara

Pune, India

Aug 2023 – Aug 2025

- Built backend systems for a conversational Taxability Research Assistant used by 20,000+ users in the US, enabling reliable LLM-assisted question answering in a high-stakes domain.
- Engineered a RAG system over 3M+ tax documents, reducing average search time from 2 minutes to 10 seconds (11x improvement).
- Productionized the service with FastAPI, Docker, Kubernetes, Prometheus, Grafana, and OpsGenie.
- Won 1st Place and Best Business Hack at Avalara Global Hackathon 2023.

Engineering Intern

Avalara

Pune, India

Jan 2023 – Jul 2023

- Collaborated on a PDF chatbot with real-time rendering and query-based highlighting, integrating a transformer-based table detection and table structure recognition pipeline through a REST API.

Data Science Intern

Coherent Corp.

Hyderabad, India

May 2022 – Jun 2022

- Delivered an ARIMA/SARIMA forecasting pipeline and dashboards for storage capacity planning and anomaly detection, decreasing prediction error by 38%.

PROJECTS

Protein-Inspired 3D Diffusion and Reward-Guided Alignment | *Python, PyTorch, Diffusion Models, RL*

- Built a 3D diffusion pipeline for synthetic protein-inspired backbones with helix/strand/coil generation, geometry-aware preprocessing, and DDPM/DDIM sampling.
- Implemented and evaluated DDPO-style reward-guided fine-tuning with a local strandness reward to steer samples toward strand-like structures, and studied stability tradeoffs using anchored reinforcement learning.

Efficient Fine-Tuning of Transformer Models Under Resource Constraints | *Python, PyTorch, PEFT*

- Benchmarked full fine-tuning, LoRA, and QLoRA on DistilBERT.
- Found LoRA reduced trainable parameters by about 98% and QLoRA reduced peak memory by about 47% versus full fine-tuning.

SKILLS

Machine Learning & GenAI: PyTorch, TensorFlow, Hugging Face, Transformers, PEFT, LoRA, QLoRA, diffusion models, flow matching, reinforcement learning, LLM evaluation

Distributed & Cloud: Docker, Kubernetes, AWS (EC2, SageMaker, S3)

Backend: FastAPI, Flask, Django, REST APIs, Celery, Microservices, GitLab CI/CD, DevOps, MLOps

Languages: Python, C++, Java, JavaScript, SQL, C