Mahek Vora

+1 412-709-2980 | mbvora@cs.cmu.edu | mahekvora.mobirisesite.com | linkedin.com/in/mahekvora | github.com/mahekvora

EDUCATION

Carnegie Mellon University — *Pittsburgh, PA*
Master of Science in Computer Science | **GPA: 4.00/4.00** — *Dec 2027*
- *Coursework: Advanced Topics in ML (PhD level), ML with Large Datasets, Distributed Systems, Deep RL*

Indian Institute of Technology Guwahati — *Guwahati, India*
B.Tech in Computer Science and Engineering | **GPA: 9.13/10.00** — *May 2024*
- *Coursework: Data Structures & Algorithms, Operating Systems, Computer Networks, ML, AI, NLP, Database Systems*

EXPERIENCE

JPMorgan Chase & Co | ***Quantitative Research Analyst*** | *Bengaluru, India* — *Jun 2024 - Jul 2026*
- Architected a **multi-agent RAG** pipeline over **Neo4j graph** and **vector embeddings** for ranked **information retrieval** and **document QnA** over model implementations, versions, and meeting minutes, cutting months of manual review
- Engineered a **Fast-API** web-app to **automate** building **query agents** over **kdb+** and **AWS Data Lake** trading databases with **multi-agent AI**, adopted across **Trading** and **Coverage** teams in **Global Markets**
- Built agent evaluation with **LangSmith** to validate **AI-generated output** for correctness and version-aware filtering

JPMorgan Chase & Co | ***Quantitative Research Intern*** | *Bengaluru, India* — *May 2023 - Jul 2023*
- **Cut runtime by 12x** by redesigning capital-calculation pipelines for stressed risk models (**VaR**, **Expected Shortfall**, **Monte Carlo**) via a sliding-window algorithm and **parallelized computation**
- **Diagnosed** data-quality gaps in risk-factor **time series** and engineered proxy-based solutions, reducing margin capital

KPMG | ***Machine Learning Intern*** | *Mumbai, India* — *Apr 2022 - Jun 2022*
- Standardized ML experiments and improved **interpretability** and **explainability** by integrating **MLflow** for model tracking and versioning across client engagements

Caliche Private Limited | ***Machine Learning Intern*** | *Guwahati, India* — *Dec 2021 - Feb 2022*
- **Designed** a two-stage **UNet computer-vision** pipeline in **PyTorch** for defect **classification** with a highlighting GUI
- Achieved **94%** validation accuracy via **ReLU** activation, **AdamW** optimization, and residual networks

RESEARCH & PUBLICATIONS

Carnegie Mellon University | *Advisor: Prof. Min Xu* — [Paper]
- **Designed and trained** an unsupervised **autoencoder** with a custom **Multi-Choice Learning** module to disentangle pose and semantic **embeddings** of protein structures, outperforming RELION by **40% ARI** on low-SNR **Cryo-ET** data
- Identified **previously unresolved ATP synthase structure** in the thylakoid membrane of yeast using this method
- **Publication (NeurIPS Poster 2026):** *DUIL: Deep Unsupervised Inverse Learning for in situ Macromolecular Morphology ID*

University of Pennsylvania | *Advisor: Prof. Vivek Gupta & Prof. Vivek Srikumar* — [Paper]
- **Co-authored TempTabQA**, an **11,545**-pair QA dataset spanning **1,208** Wikipedia infobox tables across 90+ domains, exposing a **13.5+** F1-point temporal-reasoning gap between top LLMs and human performance
- **Publication:(EMNLP Main 2023)** *TempTabQA: Temporal Question Answering for Semi-Structured Tables*

University of New South Wales | *Advisor: Prof. Rohitash Chandra* — [Paper]
- **Applied S-BERT** embeddings and **AFINN Lexicon** to quantify **NLP**-based sentiment variance across **5** Bible translations
- **Preprint (BDCC 2026):** *Large language Models for Bible sentiment analysis: Sermon on the Mount*

SELECTED PROJECTS

RL-Based Data Attribution Framework for PPO | *TRAIS Lab, University of Illinois Urbana-Champaign*
- **Designing** an **RL**-based **data attribution** framework identifying training trajectories driving **PPO** policy decisions in **time-series foundation models**, improving attribution precision by **20%** over baseline methods

Compiler for nanoC (C-Like Language) | *Implementation of Programming Languages Lab, IIT Guwahati*
- **Built a compiler** translating 3-address code to **GCC**-compatible assembly for a **C** subset (nanoC), implementing lexical analysis and parsing **algorithms** with **Flex** and **Bison**; secured an **AS** grade, ranking first in Computer Science Department

xv6 Kernel Extensions | *Operating Systems Lab, IIT Guwahati*
- **Extended xv6** kernel with **POSIX-like kernel threads**, **spinlocks/mutexes**, and a **hybrid SJF/Round-Robin scheduler**; optimized memory management via **lazy allocation** and **swap-based paging** to reduce overhead

Mini Unix Shell | *Systems Programming Lab, IIT Guwahati*
- **Implemented** a **Unix-style shell** in **C** from scratch, supporting command parsing, **process control** (fork/exec/wait), and **I/O redirection**; secured an **AA** grade, ranking first in Computer Science Department

HONORS & AWARDS

Amazon ML Summer School: Top **3.5%** of students nationally, month-long ML program
JPMorgan Quant Finance Mentorship Fellow: Top **1%** of women nationally
D. E. Shaw DESIS Ascend Educare Fellow: Top **0.5%** of women nationally as STEM mentee

SKILLS

Languages: C, C++, Python, Go, JavaScript, SQL, Bash
Technologies: Amazon EC2, Docker, AWS Data Lake, Linux
ML Systems: NumPy, Pandas, PyTorch, TensorFlow, MLflow, Keras, Scikit-learn, Matplotlib, NLTK, VLMs, CUDA, DDP
Agentic AI Systems: Model Context Protocol (MCP), LangGraph, RAG, Vector Embeddings, Neo4j, FalkorDB, HNSW, FAISS
Web & App: FastAPI, Node.js, MongoDB, Streamlit, Panel, Flutter, Dart, Android Studio