AARYAN MEHTA

adm8315@nyu.edu | (602) 497-8599 | Portfolio | LinkedIn | GitHub

EDUCATION

Master of Computer Science | New York University | GPA: 3.4/4 | New York, NY Sept 2023 – May 2025
Relevant Coursework: Artificial Intelligence, Machine Learning, Data Structures & Algorithms, Big Data, Database Management, Data Science.

Bachelor of Computer Engineering | Mumbai University | GPA: 8.7/10 | Mumbai, India Aug 2019 – May 2023

PROFESSIONAL EXPERIENCE

Institute of Fine Arts, New York University *Sept 2024 – Present*
AI/ML Student Developer *New York, NY*

- Developed an AI based video translating and multi-language subtitle generation pipeline using Python, OpenAI's Whisper model, PyTorch and Hugging Face, achieving real-time video processing in 100+ languages with 99.7% accuracy, a 200x speedup over traditional methods.
- Integrated machine learning models into a microservices architecture using Docker and Kubernetes, optimizing real-time deployment pipelines, and improving fault tolerance and model scalability in production.
- Implemented CI/CD pipelines for continuous deployment of AI models, improving deployment speeds by 30% and enabling seamless updates in production environments.
- Spearheaded a comprehensive overhaul of Dean's website using React, implementing a modular component architecture that reduced codebase complexity by 40% and improved development velocity by 3x.
- Engineered a blazing-fast, server-side rendered React application with Next.js, achieving a 300% improvement in initial page load times and boosting the website's SEO ranking by 200 positions for key search terms.

Emkay Global Financial Services Ltd. *May 2020 – May 2024*
Senior Machine Learning and SWE Intern *Mumbai, India*

- Contributed to the development of an AI trading bot as part of a team, using an ensemble of LSTM, Random Forest, CNN, SVM, and Deep Reinforcement Learning models. The AI bot was able to accurately forecast short and long-term stock movements resulting in a 25% boost in trading efficiency and generating ₹15 million in additional annual revenue.
- Achieved 30% faster market trend identification by integrating real-time Bloomberg data into an AI-driven prediction pipeline, allowing Emkay's traders to make faster, data-backed decisions.
- Reduced trading errors and losses further by 40% through integration of a reinforcement learning agent that dynamically fine-tuned predictions based on changing market conditions, directly contributing to real-time, high-stakes decision-making.
- Generated accurate market sentiment insights by developing a fine-tuned custom LLM for Twitter sentiment analysis, enabling Emkay analysts to factor in social sentiment for specific stocks a more comprehensive stock evaluation.
- Achieved near-zero downtime and real-time performance by architecting and deploying the system on AWS with Docker and Kubernetes, leveraging Kafka for high-throughput, real-time data streaming from Bloomberg's financial API, benefiting hundreds of Emkay's active traders.
- Enabled natural language query processing for hundreds of traders by creating an NLP microservice that transformed human-like queries into actionable trading insights, increasing trader engagement and client satisfaction.
- Collaborated with Emkay's CyberSec team to secure sensitive financial data with zero incidents by deploying a secure, multi-layered architecture compliant with Bloomberg data handling policies and using AWS security protocols, protecting Emkay's trading insights from data breaches.

PROJECTS

Fine-Tuning Large Language Models for Patent Classification and Semantic Search
Tech Stack: *Python, PyTorch, Hugging Face Transformers, BERT, T5, TensorFlow, Docker, Kubernetes.*

- Led the fine-tuning of advanced language models like BERT and T5 on large-scale patent datasets, achieving a 25% boost in classification accuracy and significantly improving patent similarity search.
- Designed a lightning-fast search pipeline using custom embeddings, cutting query response times by 40%, allowing users to instantly find relevant patents in a sea of legal jargon.
- Deployed the models in a fully containerized Docker environment, ensuring real-time scalability and 99.9% uptime, allowing seamless integration into existing patent search systems.
- Implemented domain-specific tokenization and tailored attention mechanisms, driving a 30% increase in search relevance while maintaining top-tier performance under heavy user traffic.
- Worked closely with patent experts and cross-functional teams, transforming their workflows and improving patent retrieval speeds by 35%, resulting in higher user satisfaction and engagement.

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, JavaScript, TypeScript, C, C++, R, Go.
- **AI/ML:** Neural Networks, Transformers, Generative AI, Large Language Models (LLMs), Transformers, Deep Reinforcement Learning, ROS, Computer Vision, NLP.
- **Cloud Platforms, DevOps & Databases:** AWS, Kubernetes, Docker, Jenkin, GIT, CloudSQL, RDS, AlloyDB, DynamoDB.
- **Frameworks and Libraries:** React, Node.js, Next.js, REST APIs, TensorFlow, PyTorch, Keras, Scikit-learn, OpenAI Gym, Hugging Face, OpenCV, SciPy, Flask, Django, HTML, CSS.