# Simardeep Singh Mehta

+1-934-222-8000 sm11377@nyu.edu linkedin.com/in/simarnyu github.com/simarmehta New York City

## Education

**New York University** **September 2023 – May 2025**
*M.S. Computer Science; GPA: 4.0/4.0* *Brooklyn, New York*
- Selected Coursework: Big Data, Cloud Computing, Computer Vision, Machine Learning, Design and Analysis of Algorithms

**Amity University, Uttar Pradesh** **July 2019 – June 2023**
*Bachelors in Computer Science; GPA: 3.77/4.0* *Noida, India*

## Experience

**American Institute for Research** **May 2024 – August 2024**
*MLOps Intern* *New York, USA*
- Architected and executed an MLOps pipeline using Flask and ChromaDB, integrating with Open source Llama model and AzureOpenAI for an ensemble Retrieval Augmented Generation (RAG) system, enhanced with re-ranking methods and performance tuning to ensure semantically meaningful document chunks are fetched, resulting in a 25% improvement in document retrieval efficiency.
- Orchestrated end-to-end CI/CD workflows with automated testing, dockerization, and Kubernetes deployments, reducing deployment times by 20%.
- Designed Helm charts for Kubernetes, facilitating version-controlled, templated deployments of a microservices architecture optimizing environments for reliability and performance.

**Ernst & Young (EY)** **June 2022 – August 2022**
*Data Science Intern* *Gurugram, India*
- Built an end-to-end ETL pipeline using AWS Glue and Lambda to process 3TB+ of financial data daily with 99.9% accuracy. Refined data by correcting outliers and handling missing values, integrating automated analysis and visualization with AWS QuickSight and SageMaker to ensure data quality compliance and monitor for data drift.
- Implemented a distributed data processing system using Apache Spark and Pyspark on EMR, reducing data processing time by 40%.
- Developed a pipeline using Python and Pandas to analyze SAP transactional data, implementing rule-based and statistical methods to detect anomalies.

**OnePercent Software LLC** **April 2022 – June 2022**
*Software Engineering Intern* *Bengaluru, India*
- Collaborated with Data Scientists and Machine Learning team to implement and scale a reinforcement learning model using TensorFlow to dynamically adjust game difficulty in a mobile game, leading to a 25% increase in user retention.
- Engineered and deployed a microservices architecture for backend using Docker and Kubernetes for the backend, improving scalability, and reducing server costs by 20%.
- Enhanced efficiency and user satisfaction by 15% through bug fixes, unit testing, feedback mechanisms, and developing a real-time analytics pipeline using the ELK stack (Elasticsearch, Logstash, Kibana) to monitor user behavior and game performance.

## Technical Skills

**Languages**: Python, C++, R, HTML/CSS, MySQL, Dart, Django, JavaScript
**Frameworks**: Pytorch, Scikit, TensorFlow, Keras, OpenCV, Flask, MongoDB, PostgreSQL, MapReduce
**Libraries**: NumPy, Pandas, Matplotlib, NLTK, HuggingFace, Spacy, Seaborn
**Tools**: AWS, GCP, Helm, Git, Terraform, Spark, Docker, Kubernetes, Linux

## Projects

**EaTexas - Food Delivery App** | *AWS Step Fucntions, Lex, SageMaker* **Project Code**
- Implemented a multi-modal machine learning pipeline using AWS Step Functions, integrating Lex with BERT for text analysis while leveraging AWS Rekognition alongside ResNet for advanced image processing of food items, for more accurate recommendations and analysis.
- Designed a real-time monitoring system using AWS CloudWatch and Grafana, setting up alerts and dashboards for admin panel of the application.
- Engineered a scalable, secure and serverless architecture using AWS Lambda, API Gateway, and SQS for a chatbot-based food delivery system, efficiently handling 1000+ concurrent users.

**FashionGPT - Product Recommendation System** | *CNN, ResNet50, VGG16, FastAI* **Project Code**
- Engineered an AI-driven fashion recommendation system combining ResNet50, VGG16, collaborative filtering, and content-based methods, leveraging FastAI for transfer learning, improving product match accuracy by 30%.
- Implemented A/B testing framework using Python and Bayesian methods to optimize recommendation algorithms, resulting in a 15% increase in click-through rates.
- Automated data collection for 100,000+ items using BeautifulSoup, cutting manual entry by 70%, and designed a feature store with Feast to enable efficient feature sharing across ML models, reducing training time by 30%.
- Deployed the system on Azure Kubernetes Service (AKS), implementing canary deployments and rollbacks and achieving 99.99% uptime.

## Honors & Awards

AWS Academy Graduate and Practitioner - AWS Academy Cloud Foundations. Issued on 09/28/2021.

*"Unraveling Information about Deep Learning."* (IRJET), Vol. 09, Issue 11, November 2022.