AIDAN SINGH

US Citizen | New York, NY | (518) 982-8994 | aidanxaxs@gmail.com | LinkedIn | Website | GitHub

EDUCATION

Cornell Tech (Cornell University), New York, NY — May 2024
Master of Engineering in Electrical and Computer Engineering
Select Coursework: ML Engineering, ML Hardware, Natural Language Processing, Digital Signal Processing

New York University, New York, NY — May 2023
Bachelor of Music in Computer Science and Music Technology

TECHNICAL SKILLS

Coding Languages: Python, Java SE, TypeScript, JavaScript, SQL, C#, C, C++, x86 Assembly, Matlab, CUDA
Tools: PyTorch, SciKit-Learn, Apache Spark, Apache Airflow, Dataflow, Pandas, Docker, Kubernetes, AWS, GCP

EXPERIENCE

Amazon Web Services, Software Development Engineer, (Java, Python, TypeScript, Kubernetes), December 2024 – Current
SDE for AWS Bedrock building generative AI inference and service infrastructure, notably hosting Anthropic's Claude Code and Chat

- Key contributor scaling an internal Bedrock data lake from launch to over 1 petabyte in 6 months, 150M rows/day
- Enabled visibility of Anthropic model hosting compute usage from disparate regional Kubernetes clusters with ETL
- Contributed to efforts eliminating monthly billing discrepancies leveraging the Databricks Medallion Architecture
- Delivered data export infrastructure and Claude resource escrow designs directly to Anthropic Technical Staff
- Enabled launch of Responsible Scaling Policy for Claude on Bedrock by implementing cross-region aggregation, load testing
- Streamlined LLM evaluation workflows by combining internal metrics packages for RAG and programmatic inference
- Delivered code leveraging AWS micro services: Lambda, Glue, SQS, SNS, Step Functions, Athena, Firehose, S3, Cloudwatch

Tulip AI, Audio AI Scientist (Python, GCP), January 2024 – May 2024

- Built an annotated cultural audio dataset with ethical web scraping leveraging the library of congress and royalty free sites
- Identified and recommended best pre-trained generative models for Tulip's use-case via contrastive analysis
- Conducted literature review and proposed fine-tuning methodology for Meta's audio-craft generative audio models
- Designed coding and data collection tasks for undergrad interns, exercised asynchronous communication & technical reviews

Universal Music Group, Data Engineering Intern (Python, SQL, GCP, Apache Airflow, Dataflow), June 2022 – August 2022

- Enabled social media analysis by delivering a production ETL pipeline ingesting music-usage data from Meta (stories, reels)
- Identified $100k/year in potential GCP savings by proposing structured database redesign and schema optimizations
- Presented GCP Cost saving proposal to VPs and assistants of CTO, leading to contractors later executing the proposal
- Collaborated effectively with another intern and engineering team using data lake exceeding 10 Petabytes

NYU Music and Audio Research Lab/ National Science Foundation, REU (Python, C/C++**),** November 2021 – May 2022

- Contributed to micarraylib, an open-source Python library for audio data aggregation for training deep learning models
- Translated audio algorithms from ambisonic encoder 'Array2SH' from C to Python with unit testing
- Participated in National Science Foundation REU Program, advised by Iran R. Roman under Juan P. Bello

PROJECTS

MiniTorch: Efficient Deep Learning Library (Python, Numba, CUDA) 2024

- Implemented all PyTorch functional abstractions in pure python: Tensors, autodiff, backpropagation, broadcasting, gradients
- Optimized code for GPU hardware using just-in-time compilation, CUDA programming (nvidia chips), custom matrix algos
- Built and trained neural networks end-to-end with library (MLPs, CNNs) and validated behavior against PyTorch

Song Genre Classification (Python, SKLearn) 2023

- Designed and trained an SKLearn AdaBoost machine learning model to classify song genre using features from Spotify's API
- Achieved an average AUROC score of .88 for classifying 10 musical genres
- Prepared data from 50,000 songs through feature extraction and handling missing and low-quality data

Spatial Drawing Application (C#, Unity) *Code available upon request* 2023

- Created a unity based spatial design application (.apk) for VR, developed on Meta Quest
- Implemented locomotion and object manipulation from scratch utilizing the Unity XR Interaction Toolkit and 3D math
- Spearheaded object placement (placing, scaling and orienting primitive objects) within a group project