Benjamin E. Jordan

bej9@cornell.edu | 607-339-1740
LinkedIn | Personal Website

Skills

Languages: Python, Java, TypeScript, C++, C **Other**: Git, Docker, AWS, Runpod, Linux, HuggingFace, LaTeX

Tools/Libraries: NumPy, Spark, Pandas, Pytest, PyTorch, Scikit-learn, Boto3, Faiss, DDP/FSDP, Optuna, Angular, Selenium

Experience

Software Engineer @ Amazon Oct 2024 – Present
Measurement, Ad Tech, & Data Science *New York, NY*

- Productionized brand-awareness prediction model pipeline, which processes 30+ million rows of data daily, aggregating survey responses and model predictions to produce metrics for brand lift reports
- Designed and delivered a workflow to automate the model training, evaluation, and approval process on a weekly cadence
- Created a small design to enable shadowing of experimental models in a production environment
- Worked extensively on large-scale spark ETL jobs (in AWS EMR and AWS Glue)
- Also contributed to various other components owned by team in Java, Python, Typescript, and Scala
- Collaborated frequently with surveys science team to implement new methodologies for measuring advertising performance

Machine Learning Engineer @ Northrop Grumman Feb 2024 – Oct 2024
Autonomous Intelligence & Robotics Lab *Denver, CO*

- Enabled accurate, real-time object detection by distilling GroundingDINO into YOLOv10, decreasing inference latency on Jetson Orin from 0.6s to 0.08s
- Researched techniques for implementing and evaluating retrieval-augmented generation (RAG) systems, and led the technical side of a project focused on classified PDF retrieval and question answering
- Implemented a RAG system utilizing ANN search and locally running vision-language models for PDF Q&A
- Wrote a survey on techniques for mitigating a lack of labeled training data: active learning, semi-supervised learning, knowledge distillation, and synthetic data generation

Machine Learning Engineer @ KLA May 2023 – Aug 2023
Internship *Ann Arbor, MI*

- Independently worked on semiconductor defect detection algorithms
- Fine-tuned and quantized vision transformers for efficient and robust defect classification with limited data
- Presented project during poster board session and was invited to give a virtual talk on transformers and transfer learning

Software Engineer @ Carestream May 2022 – Aug 2022
Internship *Rochester, NY*

- Developed and maintained backend functionality in Carestream's ImageView x-ray software

Projects

Entropy Audio | **Project Blog** | **Website** Jan 2024 – Present
Personal project focused on novel sound generation tools for composers

- Curated a multi-terabyte audio sample dataset using web scraping, open-source data, synthetic description generation and augmentation, and manual labeling
- Created code packages for training, metrics, model code, and data, drawing inspiration from open source codebases: stable-audio-tools (Stability AI), audiobox-aesthetics (Meta FAIR), audiocraft (Meta FAIR)
- Trained a 2B parameter audio diffusion model on audio sample dataset
- Implemented algorithm from the Diffusion-DPO paper and performed post-training using a generated preference dataset, self-collected through the EntropyAudio website UI
- Implemented a frontend app for text-based audio generation that incorporates a preference data flywheel into the workflow
- Implemented a backend system in AWS to support generation, data collection, and other application logic
- Created a serverless inference endpoint on Runpod which gets called by a lambda in AWS

Education

Cornell University 2022 – 2023
Master's in Computer Science, Concentration in Machine Learning, 3.76 GPA *Ithaca, NY*

Rochester Institute of Technology 2018 – 2022
Bachelor's in Computer Science, Minor in Music, Magna Cum Laude, Presidential Merit Scholarship *Rochester, NY*

Notable Coursework

Machine Learning: Large-Scale Machine Learning, Machine Learning Theory, Reinforcement Learning, Machine Learning Hardware and Systems, Computer Vision, Numerical Linear Algebra, Natural Language Processing,

Systems: Distributed Computing, Parallel Computing, Computer Architecture, Operating Systems, Networks