

# ERIC ZACHARIA

## ARTIFICIAL INTELLIGENCE SCIENTIST

### WORK HISTORY

#### Senior Machine Learning Engineer - *QuayChain*

Specialized in the creation, optimization, and deployment of custom machine learning, deep learning, and computer vision models as the sole Machine Learning Engineer for an early-stage start-up SaaS business. I worked very closely with our Data Engineer in developing our AWS pipeline to solve multi-camera based problems that made predictions about traffic flow. I analyzed raw video data using image processing, computational photography, machine learning, and deep learning. Our product was a web application that supply chain managers would use to help them validate the arrival of their trucks, chassis, and containers at ports around the country. The AI powered dashboard provided location and time-stamped images of trucks along with relevant information such as the license, USDOT, chassis id, and container numbers, which served as an automated proof of delivery.

1/2022 to  
3/2023

#### Teaching Assistant - *University of Chicago*

Assist with teaching and grading for UChicago's Master's Program in Computer Science.

9/2021 to  
present

#### Machine Learning Engineer - *UChicago Medicine*

Developed machine learning models trained on the hospital's database of patient data to help suggest proper ICD codes for doctors and flag patients with a high risk of cardiovascular diseases.

1/2021 to  
1/2022

#### Machine Learning Research Intern - *Argonne National Laboratory*

Worked with molecular engineers and utilized GPU-powered Bayesian inference to compensate for drifting pH and voltage measurements in water contamination sensors.

6/2020 to  
9/2020

#### Senior Aerospace Stress Engineer - *Spirit AeroSystems*

Worked on small engineering teams that designed, built, analyzed, tested, and repaired Boeing's 777X and 787 Dreamliner, and a secret aerospace structure for the U.S. DOD.

6/2016 to  
5/2020

#### Fluid Dynamics Researcher - *Green Fluids Laboratory*

Analyzed dolphins and experimented with 3D-printed fins that swam inside a water tunnel. Stitched together 2D 4K videos into 3D vortex visualizations for further fluid analysis. Recipient of George M. Berry Award for Outstanding Design Achievement in Engineering.

5/2015 to  
5/2016

### PROJECT EXPERIENCE

#### Port Logistics Classifier - Computer Vision Project at *QuayChain*

Summary: Collection, Labeling, Augmentation, Training, Detection, Tracking, Clustering, OCR, Aggregation, and Deployment

Took an early-stage supply-chain automation project from data collection to deployment using live recordings of congested ports in CA and UT.

- Ffmpeg-split videos of traffic at ports into images to serve as data for my object detection model.
- Initially hand-labeled frames with multi-class bounding boxes to classify license plates, trucks, chassis, containers, hazard placards, and cars.
- Trained a small YOLOv8n model to assist with data labeling, enlarged the dataset with augmented copies, and trained a robust model for production.
- Detected small distant license plates in 4K videos using a tiling method that splits the image into smaller images during training and inference.
- Optimized training time with intelligently selected batch sizes and EC2 instances to handle the memory requirements for 4K resolution training.
- Solved multi-lane vehicle occlusion challenges by ensembling the YOLO model with custom PyTorch feature extraction and StrongSORT tracking.
- Fortified license plate detection with a real-time Canny edge detection algorithm that finds contours with aspect ratios similar to license plates.
- Used a combination of OCR and unsupervised clustering models to get the texts-of-interest amongst the abundance of other text shown on the trucks.
- Aggregated trucks using the port between variable datetimes, then Dockerized and deployed an interactive traffic stats dashboard for truck managers.
- Triggered uploading of passing trucks to the Cloud using detection and used Lambda to pass frames through the pipeline for extracting relevant text.
- Pipeline results funnel to another online dashboard that shows managers a truck image with the datetime, location, license, container numbers, etc.

#### ICD Code Suggester and Cardiovascular Disease Flagger - Natural Language Processing Project at *UChicago Medicine*

- Data wrangled labeled medical notes from various CSV files into dataframes using Pandas and ngram methods from nltk.
- Each data sample corresponded to a single patient, and column features included medical notes, sex, medication history, lab results, etc.
- Labels included correct ICD codes, which doctors use to diagnose patients, and three different cardiovascular diseases that patients may have had.
- Medical notes were vectorized into sparse n-gram feature columns, and categorical features were numbered before feeding into XGBoost.
- Predictions were mediocre due to n-gram sparsity, so I hoped to improve the context gained from medical notes using BERT's attention mechanism.
- BERT significantly outperformed the n-gram XGBoost model, hinting that BERT had unveiled more predictive context from the medical notes.

#### Predicting the Genre of Music using a Convolutional Neural Network - Click [here](#) to see my video presentation

- Used audio signal processing techniques to transform MP3 files into Mel spectrogram images that served as genre labeled training data
- Wrote a CNN model in Keras to see if it could outperform benchmark machine learning classifiers trained on extracted audio features. It did!
- Developed a Streamlit web application for users to upload MP3 files for genre classification; it displays the Mel spectrogram and genre prediction.

#### Predicting the Stock Market with Sentiment Analysis of Live Tweets - Click [here](#) to see my video presentation

- Developed an NLP pipeline that algorithmically trades stocks using the sentiment of tweets about a portfolio of stocks from Twitter
- The trading bot successfully decided to purchase more of the stocks with bullish sentiment and avoided buying those with bearish sentiment.
- The trading pipeline was built using PyTorch, BERT, the Twitter Developer API, and a paper trading API from Alpaca.

#### Compensating for Drifting Measurements in Water Sensors - Probabilistic Programming Project at *Argonne National Laboratory*

- Sensor drift is a natural phenomenon that occurs in all sensors caused by the physical changes in the sensor.
- Plots measuring pH and voltage vs. time took the form of exponential distributions that steadily decreased during asymptotic approach.
- Implemented Bayesian Inference, using an exponential prior distribution, random sampling of points with PyMC3, and computation using Colab.
- Successfully generated compensated plots using the posterior distribution, drawn from the likelihoods of sampled points given the experimental data.

### CONTACT

[UnsupervisedBias.com](https://www.unsupervisedbias.com)

[GitHub](#) | [LinkedIn](#) | [Web Portfolio](#)

[ekzacharia@gmail.com](mailto:ekzacharia@gmail.com) | [\(774\) 239-5342](tel:(774)239-5342)

### EDUCATION

#### University of Chicago

Master of Science, 2021

Computer Science & Data Analysis

#### Syracuse University

Bachelor of Science, 2016

Major: Aerospace Engineering

Minor: Computer Science

### SKILLS

Python, Linux, Docker, AWS (ec2, ecs, s3, lambda, sagemaker), NumPy, Pandas, Jupyter, OpenCV, Pillow, YOLO, StrongSORT, ByteTrack, ffmpeg, Git, Bash, Shell script, PyTorch, TensorFlow, Hugging Face, Sci-kit Learn, Nltk, Matplotlib, Seaborn, Streamlit, Flask, Colab, SQL, Go, JavaScript, HTML, CSS, Bootstrap, Windows, Mac, Training with M2 Max, Multi-threading/processing, AGILE

*Learning: C++, GCP, Kubernetes, PySpark*

### LICENSES

Secret Clearance, 2020

Private Pilot License, 2019

Scuba Diving License, 2016