David Zhan

301-323-5579 | dazhan@seas.upenn.edu | LinkedIn | Github

Education

University of Pennsylvania May 2026

Masters of Science in Computer Science, Bachelors of Science in Computer Science *GPA: 4.0/4.0*

- **Minors**: Mathematics, Data Science
- **Coursework:** Data Structures and Algorithms, Data Systems, Operating Systems, Machine Learning/AI, Natural Language Processing, Computer Vision
- **Involvements** Penn Aerial Robotics (Software Lead), Machine Learning Research @ Penn, Penn Club Soccer, Penn Club Badminton, Penn Flux

Professional Experience

Software Team Lead August 2024 – present

Penn Aerial Robotics — Python, ROS2, PX4, OpenCV, C++ *Philadelphia, PA*

- Engineered a computer vision payload detection system for UAVs, utilizing OpenCV for binary thresholding
- Built a custom data pipeline for real-time drone camera data processing using ROS2, enabling efficient storage, transformation, and model inference on position detection.
- Leveraged RANSAC for robust outlier detection and internal position estimation to support high-accuracy UAV navigation for SAE Aero Design 2025.

Software Engineering Intern June – August 2024

Ventoscity — Flutter, Express, TypeScript, SQLite, Python, TensorFlow *College Park, MD*

- Developed a data-driven food recognition system by integrating an AI-powered scanner achieving 90% accuracy for portion sizes and fiber content analysis.
- Redesigned and optimized the app's data storage layer, reducing SQLite query times by 30% and enabling faster retrieval of user data.
- Conducted statistical analysis of usability data from 100+ user tests, implementing 25+ enhancements that reduced app crashes by 15%.

Frontend Software Engineer March – May 2024

keep.id — React, Node.js, Express, SQL *Remote*

- Optimized SQL queries for large-scale data retrieval, leading to a 23% improvement in average document load times.
- Analyzed user interaction data from web tracking tools, identifying critical friction points that guided the redesign of the landing page, reducing user-reported issues by 60%.

Research Intern June – August 2023

National Research Foundation — Swift, SwiftUI, C++, Arduino *College Park, MD*

- Developed a full-stack mobile app for an emerging gut-health startup at UMD, playing a pivotal role in securing over $1M in venture capital funding.

Projects

Fine-tuned LSTM Sentiment Analysis Model — *numpy, pandas, scikit-learn* December 2024

- Trained an LSTM regressor on **336,239** rows of training data for ordinal classification of RateMyProfessor Reviews
- Performed data augmentation using synonym replacement and back-translation to achieve a **45%** longer data set
- Achieved a **12%** increase in Quadratic Weighted Kappa score, **15%** decrease average error in relative to pre-finetuning

Data Labeler — *React, Tailwind, FastAPI, Supabase, MEGA, AWS, Render* December 2024

- Dense captioning service used for creating high quality image training data sets
- Implemented custom quality control, aggregation metrics; Amazon MechTurk for crowdsourcing data and crowd payment
- Captioned **400+** images, with a **210%** increase in words and **50%** in relevance compared to LLM generated captions

Spotify Song Popularity Predictor — *numpy, pandas, wandb, scikit-learn, seaborn* November 2024

- Performed exploratory data analysis, data cleaning, and feature encoding on over **114k** rows of a Spotify tracks dataset
- Performed Priciple Component Analysis, Random Forest Regression, and hyper-parameter tuning, decreasing total model error by **10%**

Part-Of-Speech Tagger — *Python, numpy* September 2024

- Hidden Markov Model POS tagging implementation using **Viterbi, Beam Search and Greedy** inference methods
- Using English Penn Treebank dataset for training; achieved a **96.3%** accuracy on randomized test set

Awards

PennApps XXV: Best Entertainment Hack 2024

National Merit Scholar 2021

Skills

Languages: Python, Java, C++, Swift, Dart, JavaScript, Kotlin, Go

Frameworks: React, NodeJS, NextJS, Tailwind, TypeScript, ThreeJS, BootStrap, Flutter, Express

Tools: Supabase, Firebase, AWS, Convex, MongoDB, SQLite

Data Science/ML: pandas, numpy, scikit-learn, TensorFlow, HuggingFace, PyTorch, OpenCV