Maxim Dokukin

Able to take in a lot of chaos and turn it into something manageable

maxdokukin@icloud.com | maxdokukin.com | github.com/maxdokukin | linkedin.com/in/maxdokukin

Skills

ML Engineering: Python, C++, TensorFlow, TFLite, Classification, CNNs, LSTMs, Audio DSP, MFCCs, Confusion Matrices

Data Pipelines: NumPy, Pandas, SQL, Data Cleaning, Data Labeling, Feature Engineering, Data Provenance, ETL Scripts, Linux, Visualization

MLOps: CI/CD, GitHub Actions, Docker, GCP, Cloud Run, Model Versioning, Experiment Tracking, A/B Testing, Testing Automation, Documentation

Performance: Quantization, Latency Tuning, Memory Optimization, Profiling, Tensor Arena, DMA, I2S, I2C, Thread Priorities, UI Optimization

Education

Masters of Science, Artificial Intelligence, San Jose State University — May 2027

Bachelor of Science, Data Science, San Jose State University — May 2025

- Summa Cum Laude, Presidents Scholar, AS Leadership Scholar

Experience

Embedded ML Engineer, Nuvoton (San Jose, CA) — Aug 2024 - Present

- Built scalable **data pipelines** ensuring **data quality** and **data lineage** for **reproducibility** using **Python**, **NumPy**, **Pandas**
- Designed and deployed **TensorFlow**/**TensorFlow Lite audio classification** models, cutting RAM 21% via **quantization**
- Implemented **C++ production-grade inference** architecture meeting **latency** targets, enabling **multi-model deployment**
- Automated **experiments** and **offline evaluation**; scheduled batch runs with standardized **metrics** and **confusion matrices**
- Established **model lineage** and **documentation**: model_id, config emission, **CI/CD** with **GitHub Actions**

Machine Learning Engineer Intern, Nuvoton (San Jose, CA) — May 2024 - Aug 2024

- Boosted on-device **classification** accuracy 49%→90% using **Python**, **TensorFlow**, rigorous **evaluation**
- Built YAML-driven **data pipeline** with checksums ensuring **reproducibility**, **data quality**, and **lineage**
- Applied int8 **quantization** and memory profiling to optimize **inference latency**, footprint, and **reliability**
- Automated end-to-end **CI/CD** from preprocessing to on-device tests using **GitHub Actions**
- Engineered **C++** harness and **Python** evaluator; standardized **metrics** and nightly **monitoring**

Machine Learning Intern, Yandex (Remote) — May 2023 - Aug 2023

- Automated **Python data pipelines** with **Pandas**/**NumPy**, processing 1M+ records/320GB daily on **Linux**
- Unified preprocessing for train/validation via **CI/CD**, ensuring **data quality**, **reproducibility**, restoring performance
- Developed dashboards exposing sampling bias; enabled **model monitoring** for **data drift** during **NLP** training
- Implemented **automation** for remote servers with Bash/Cron; reliable **Linux batch jobs** and logging
- Authored standardized preprocessing **documentation**; led cross-functional workshop, improving collaboration, clarity, and **reproducibility**

On-Campus Involvement

President, AI & ML Club @ SJSU — Jan 2025 - Present

Officer — Dec 2023 - Jan 2025

- Increased club attendance from 25 to 85 members
- Expanded leadership team from 5 to 27 Officers
- Collaborated with 2 industry leaders and organized on-campus speaker-events for them
- Lead club projects with **reinforcement learning**, **neural networks**, **classification**, **clustering**, **regression models**, and **model optimization** using **TensorFlow** and **PyTorch**

Projects

Audio Data - Training

- Built end-to-end audio **classification** pipeline using **Python**, **NumPy**, **TensorFlow Convolutional Neural Networks**, achieving 85% accuracy
- Converted trained model to **TensorFlow Lite**; applied **Model Quantization** for optimized edge **inference** latency
- Integrated autogenerated **C++** model file into deployment; implemented double-buffered **real-time** audio streaming
- Designed sliding-window **Data Pipelines** with **MFCC/Spectrograms**; improved signal utilization and model robustness
- Implemented high-level **inference** filter, confidence thresholds, tie-resolution; improved prediction reliability under noise

Audio Data - Auto Testing

- Built automated end-to-end **data pipelines** for edge **classification** using **Python**, integrating **C++**
- Implemented end-to-end **automation** for repeated runs; enabled controlled **A/B testing** across versions
- Computed confusion matrix, **precision**, **recall**, **F1** with **Pandas**; standardized **offline** evaluation
- Captured edge predictions via serial; ensured **data lineage**, **reproducibility**, and audit-friendly logs
- Delivered production-grade **Python**/PowerShell scripts; cut manual testing time and errors through **automation**

Personal Portfolio Website

- Shipped LLM-powered portfolio assistant, end-to-end **Python** backend enabling conversational **LLMs** inference
- Containerized service with **Docker**, deployed on **GCP Cloud Run** via automated **CI/CD**
- Implemented SSL, **security**, and **privacy** safeguards using Cloudflare; followed **Responsible AI** practices
- Designed persistent data layer using **SQL** (MySQL), enabling reliable sessions, context-aware interactions
- Achieved low **latency** real-time chat via WebSockets/SocketIO; streamlined inference request handling